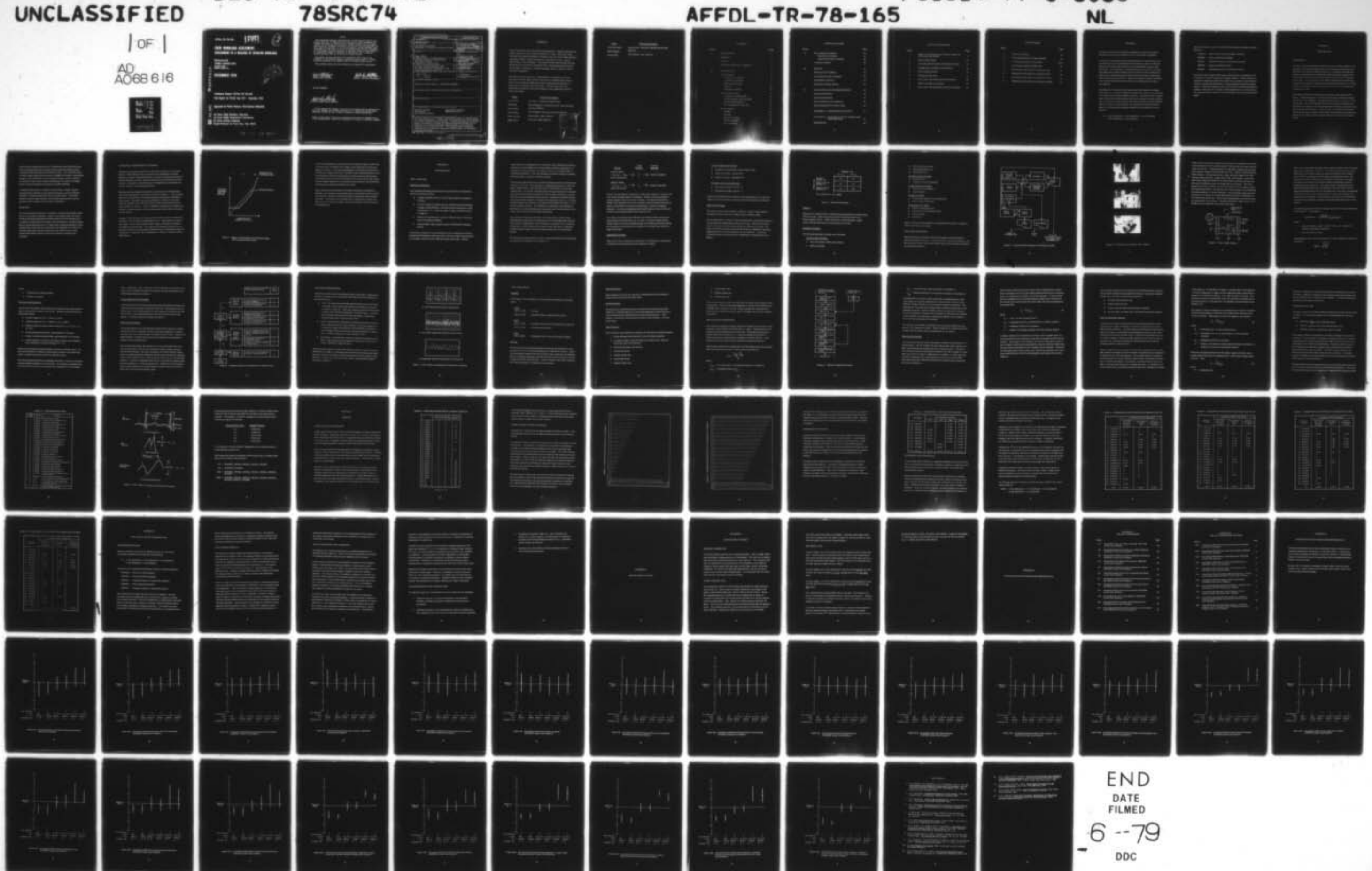
AD-A068 616 HONEYWELL INC MINNEAPOLIS MINN SYSTEMS AND RESEARCH --ETC F/G 1/2
CREW WORKLOAD ASSESSMENT. DEVELOPMENT OF A MEASURE OF OPERATOR --ETC(U)
DEC 78 J D WOLF F33615-77-C-3065

UNCLASSIFIED 78SRC74 AFFDL-TR-78-165 NL

1 OF 1
AD
A068 616



END
DATE
FILMED
6 -79
DDC

AFFDL-TR-78-165

LEVEL

CREW WORKLOAD ASSESSMENT

DEVELOPMENT OF A MEASURE OF OPERATOR WORKLOAD

Honeywell
SYSTEMS & RESEARCH CENTER
2600 RIDGWAY PARKWAY
MINNEAPOLIS, MINNESOTA 55413

AD A068616

DECEMBER 1978

Technical Report AFFDL-TR-78-165

Final Report for Period June 1977 - September 1978

Approved for Public Release; Distribution Unlimited

Air Force Flight Dynamics Laboratory
Air Force Wright Aeronautical Laboratories
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio 45433

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Information Office (IO) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

Larry C. Butterbaugh
Human Factors Engineer

John D. Vrettos, Maj, USAF
Chief, Crew Sys Development Br

FOR THE COMMANDER

James R. Stanley, Col, USAF
Chief, Flight Control Division

"If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AFFDL/FGR, W-PAFB, OH 45433 to help us maintain a current mailing list."

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFFDL-TR-78-165	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): CREW WORKLOAD ASSESSMENT. Development of a Measure of Operator Workload.	5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report, 15 Jun[redacted] [redacted]77 ~~to~~ 1 Sep[redacted] 78
	6. PERFORMING ORG. REPORT NUMBER: 78SRC74
7. AUTHOR(s): James D. Wolf	8. CONTRACT OR GRANT NUMBER(s): F33615-77-C-3065
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Honeywell Systems and Research Center, 2600 Ridgway Parkway, Minneapolis, Minnesota 55413	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Project 2403, Task 240304, Work Unit 24030406
11. CONTROLLING OFFICE NAME AND ADDRESS: AF Flight Dynamics Laboratory (FGR), AF Wright Aeronautical Laboratories, AFSC, Wright-Patterson AFB, Ohio 45433	12. REPORT DATE: December 1978
	13. NUMBER OF PAGES: 94
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 93p.	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Workload assessment
Crew workload
Crew station evaluation
Physiological variables
Regression analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The study objective was to develop a quantitative measure of workload useful in crewstation evaluation. Flight tasks of varying difficulty were simulated, and 35 pilot response variables analyzed. Selected physiological and visual response variables were applied in a stepwise regression procedure to the prediction of a composite performance/opinion meausre, which reflected differing levels of task difficulty. The resulting linear equation was reformulated as a preliminary operationally-defined measure of information-processing workload.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

402349

FOREWORD

Work reported here was performed by Honeywell Inc., Systems and Research Center, under Air Force Contract F33615-77-C-3065, Project No. 2403. The study was monitored under the direction of the Air Force Flight Dynamics Laboratory (FGR), Wright-Patterson Air Force Base, by Mr. Larry C. Butterbaugh, Project Engineer. This report covers work performed during the period 15 June 1977 to 1 September 1978, under Honeywell direction of Dr. Stirling P. Stackhouse, Program Manager, and Mr. James D. Wolf, Principal Investigator.

The author wishes to thank Mr. Butterbaugh for his guidance and active participation in this study, and Mr. Sam Herron of Bunker-Ramo Corporation, who provided invaluable pilot factors support during preliminary evaluations of simulator task alternatives. The following pilots and Honeywell personnel are also acknowledged for their significant contributions to the effort.

Pilots	Honeywell Personnel
Lyle Andvik	Jim Datta--simulation programming
Tom Biele	Kurt Graffunder--simulation set-up, data collection, and data reduction
Dick Delano	
Lou Kressin	Terry Hanson--data collection and reduction
Pete Lovering	Brian Nelson--data reduction
Roger Lund	Bob North--data reduction

Pilots	Honeywell Personnel
Field McConnell	Doug Perlick--analysis programming and data reduction
Mike Nesgoda	John Sandvig--data collection
Tom Schulte	

CONTENTS

CONTENTS (concluded)

LIST OF ILLUSTRATIONS

LIST OF TABLES

SUMMARY

The study objective was to develop a quantitative measure of operator information processing workload for use in crew-station evaluation.

A conceptual relationship between task performance, task difficulty, and operator workload was formulated which predicts a positive correlation between performance and workload over an intermediate range of task difficulties. Test conditions of varying difficulty anticipated to be within this range were produced through use of a composite task including both flight control (simulated landing approach) and a secondary loading task (Sternberg fixed-set procedure). Data on 35 operator-response variables were collected under these test conditions with eight pilots serving as test subjects.

Resulting data for selected physiological and visual response variables were applied in a stepwise regression-analysis procedure to the prediction of a composite performance/opinion measure. The purpose was to identify linear combinations of physiological and visual response variables yielding high correlations with task performance and pilot opinion of task difficulty. Based on results of these analyses, the following operationally-defined metric for information processing workload (W) was tentatively recommended:

$$W = 0.631\,(\mathrm{EMGAAM}) + 0.103\,(\mathrm{RESPAM}) + 0.163\,(\mathrm{RESPAS}) - 0.386\,(\mathrm{RESPDM}) + 0.167\,(\mathrm{RESPDS})$$

where the mnemonics are all normalized physiological response variables defined as

EMGAAM = mean forearm electromyogram amplitude

RESPAM = mean respiration amplitude

RESPAS = standard deviation of respiration amplitude

RESPDM = mean respiration duration

RESPDS = standard deviation of respiration duration

This metric has at least ordinal scale characteristics, and therefore can be applied for relative comparison of design options. Further analysis is required to refine this preliminary metric into an interval-scale estimate capable of quantifying differences in workload demand imposed by design options. Additional validation work is also needed to demonstrate generalizability of this metric (or further refined alternative) to workload estimation associated with a variety of real-world flight and mission management tasks.

SECTION I

INTRODUCTION

BACKGROUND

Of major importance during the design of Air Force weapon systems is the capability of the pilot to interact with aircraft systems to effectively and efficiently accomplish the defined mission. It is common knowledge that mission accomplishment is significantly influenced by the complex interactions between the aircraft's control system, the information displayed, and the pilot. For this reason, crew station evaluation procedures need to address the problem of determining if presented information is provided in such a way that the pilot can best interact with the control system and accomplish the task.

Both quality and quantity of displayed information affect piloting performance. Cluttered displays, or displays which lead to inefficient competition for the pilot's attention, can degrade performance as well as too little information. Also, the means by which a given amount of information is presented with respect to the control authority of the pilot-control loop also affects piloting performance. In other words, the amount of information presented, the way in which the information is presented, and its appropriateness to the control system and task, influence the merit of an avionic system.

Other mission-related tasks such as communication and mode switching, and environmental factors including visibility and wind turbulence also contribute to the total task load imposed on a pilot. The composite effect of these system-design and operational variables can produce a measurable change in pilot performance, and an apparent change in pilot workload. An objective in any avionic system development is to achieve acceptable levels of both pilot/system performance and pilot workload.

To determine the merit of a system in these terms, a metric must be developed which reflects not only task performance, but also expresses difficulty or workload associated with task performance. This metric should be valid for both simulator and in-flight applications to have the greatest payoff in support of crew-station evaluations.

OBJECTIVE

The overall program objective is to develop a practical empirically-based tool for crew-station evaluation. Effort in the present study was concentrated on developing alternative workload metrics for this purpose based on analysis of physiological-response, task-performance, and opinion data. Specific study objectives were to 1) select test conditions expected to differ substantially in task difficulty and performance attainable, 2) collect pilot response data under these test conditions, and 3) derive one or more operationally-defined metrics which can be applied to quantify pilot workload.

CONCEPTUAL FRAMEWORK OF APPROACH

Information processing capabilities of the human operator are inherently limited. Within these limits, the operator can compensate to varying degrees for system design deficiencies or adverse operating environments. As task demands increase, operator information processing limits will be reached or exceeded, and performance will degrade at an increasing rate on one, several, or all assigned functions.

Hypothesized interrelationships between performance and workload with increasing task demand or difficulty are shown in Figure 1. This conceptualization is simplified by linearizing segments of functions to distinguish three stages. Information processing workload is assumed to increase with task difficulty only to the point where a limit on processing capacity of the operator is reached. Thus, "workload" as defined here relates most directly to actual utilization of processing capacity rather than to demands on this capacity.

In stage I, task demands are sufficiently low to allow operator compensation with little or no increase in performance error. Task demands in stage II are higher, and increased operator effort cannot completely compensate for additional task demands. The result is increasing performance error. Operator information processing (workload) limits are exceeded in stage III, producing an accelerated rate of degradation or failure in operator/system performance.

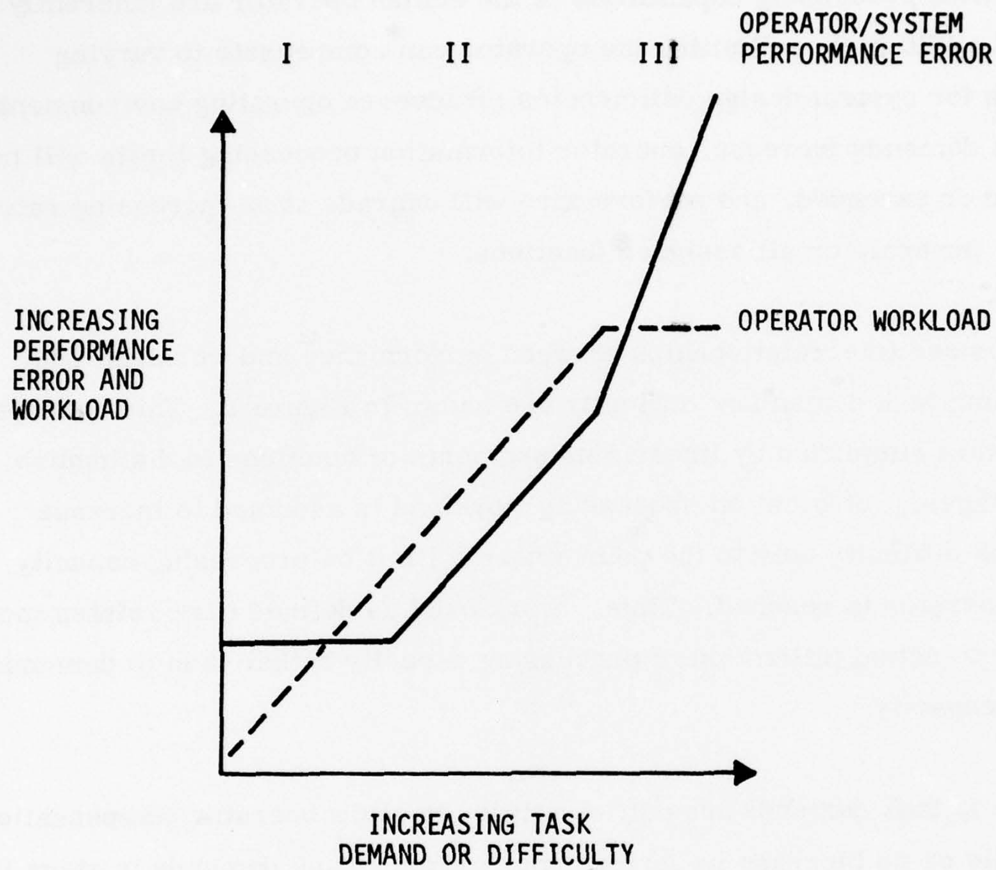


Figure 1. Stages of Performance and Workload Change with Increasing Task Demand

In this conceptualization, performance and workload measures would tend to be positively correlated over a range of task difficulties indicated as stage II in Figure 1. Previous investigations of piloting tasks selected to be within this range (References 1 through 4) have indicated positive correlations between operator physiological responses and task performance. Correlations have also been found between eye-motion activity and Cooper-Harper ratings (Reference 7), and between pupil diameter and information processing task difficulty (References 8 and 9).

Basic measurement and analysis techniques developed in References 1 through 4 were applied in the present study to an expanded variety of operator response measures including both physiological and visual response variables. The purpose was to define linear combinations of these variables which correlate most highly with task performance and pilot opinion of task difficulty.

SECTION II

METHODOLOGY

TEST CONDITIONS

Rationale for Selection

Test conditions defining piloting tasks to be performed were selected to satisfy the following criteria:

- Of basic interest to the Air Force Flight Dynamics Laboratory (AFFDL)
- A priori reasons to believe that the conditions would encompass a range of task difficulty, and would produce differences in task performance (i.e., within the stage II range of difficulty shown in Figure 1)
- Conducive to generating a variety of different types of response data for subsequent analysis
- Implementable within program scope on Honeywell's simulator facility

Preliminary evaluations of task alternatives were conducted with AFFDL personnel participating to select the type of primary (flight control) task to be simulated, and the specific difficulty levels of this task. General

requirements were established for a secondary task utilizing the Sternberg fixed-set procedure (Reference 5) with auditory stimulus presentation. A secondary task was applied in this study to obtain an increased range of overall task difficulty. Thus, the Sternberg procedure served as a loading task rather than a measure of reserve capacity.

The primary task selected was a simulated landing approach on instruments with an F-4 aircraft. This simulation, originally developed for a previous project (Reference 6), required only minor modifications for use in the current study. An AFFDL-designated pilot flew approximately 90 simulated approaches under differing system and environmental conditions to provide the basis for selecting task difficulty levels. Conditions evaluated included varying gust levels, number of approach path segments (one, two, or three), throttle (manual vs. automatic), and flight control system (nominal vs. rate limited control surfaces). The feasibility of presenting visual Sternberg-task stimuli was also evaluated and rejected because of high demands on the visual channel imposed by the flight-control task.

Three levels of flight task difficulty were defined which yielded clearly distinguishable differences in performance errors and judged task difficulty. These levels, defined below, were formed by a composite of two variables; (1) gust level, and (2) flight control system mode--nominal vs. degraded. Conditions held constant include one-segment approach path and manual throttle.

The secondary loading task was a choice reaction task based on the following procedural model (adapted from Reference 5).

Stimuli		Test Stimulus		Correct Response
Positive subset $= X_1, X_2, \ldots, X_M$ (list to memorize)	$\longrightarrow$	X_i	$\longrightarrow$	Positive response
Negative subset $= Y_1, Y_2, \ldots, Y_N$ (items not on list)	$\longrightarrow$	Y_j	$\longrightarrow$	Negative response

Briefly, the test subject is required to memorize a subset or "memory set" of M items (typically letters or numbers). Stimuli are then presented in random sequence from a set of M + N items. The subject must decide whether each stimulus is or is not a member of the M items in the positive subset, and indicate his decision by generating a positive or negative response. Previous investigations have found response times on this type of task to be an approximately linear function of M.

Two levels of secondary-task difficulty were defined without preliminary experimentation by selection of M = 2 and M = 4 positive subset stimuli. Larger values of M were excluded because of the additional familiarization and training time anticipated to be required for thorough memorization of longer item lists.

Independent Variables

Based on the above findings and constraints, the following two independent variables and associated levels were defined for testing.

Primary (flight control) Task

- No gusts; nominal flight control system (FCS)
- Gusts to 18 knots; nominal FCS
- Gusts to 30 knots; degraded FCS

Secondary (choice reaction) Task

- Sternberg procedure with M = 2
- Sternberg procedure with M = 4

Nominal and degraded FCS conditions refer respectively to 0.5 and 0.05 radian per second rate limits placed on aileron and stabilator deflection.

Experimental Design

The experimental design is shown in Figure 2. Each of eight subjects completed 10 trials per cell, yielding a total of 480 test trials.

As described below, one form of data collected was a comparative judgment of task difficulty obtained after each pair of trials. There are 30 permutations (trial pairs) of the six test conditions, taken two at a time. The total of 60 trials per subject provided the means for obtaining comparative judgment data for all 30 permutations of test conditions. Trial pairs constituting these permutations were independently randomized for each subject.

PRIMARY TASK	SECONDARY TASK: M = 2	SECONDARY TASK: M = 4
NO GUSTS NOMINAL FCS	(1)*	(4)
GUSTS TO 18 KTS. NOMINAL FCS	(2)	(5)
GUSTS TO 30 KTS. DEGRADED FCS	(3)	(6)

*TEST CONDITION OR CELL NUMBER

Figure 2. Experimental Design

Subjects

Subjects were obtained from Air National Guard squadrons currently flying RF-4B aircraft. Seven were pilots and the remaining subject was a weapon-system officer. Flight experience of these personnel ranged between 950 and 4050 hours in various types of aircraft.

Dependent Variables

The following dependent variables were recorded:

Primary Task Variables

- Root mean square (RMS) pitch attitude
- RMS roll attitude

- RMS vertical path error
- RMS lateral path error
- RMS speed error

Secondary Task Variables

- Response time
- Percent of correct responses

Visual Response Variables

- Fixation x, y coordinates
- Pupil diameter

Opinion Variables

- Comparative judgment of task difficulty
- Scalar rating of difficulty

Physiological Variables

- Electrocardiogram (ECG)
- Forehead electromyogram (EMG)
- Forearm EMG
- Respiration

Detail on definition and method of recording these variables is included in the test item description below.

TEST ITEM DESCRIPTION

Major functional components of the task simulation and data collection facility are shown in Figure 3. Layout of pilot's controls and displays, and the oculometer electro-optical unit used to sense visual response variables, are depicted in Figure 4.

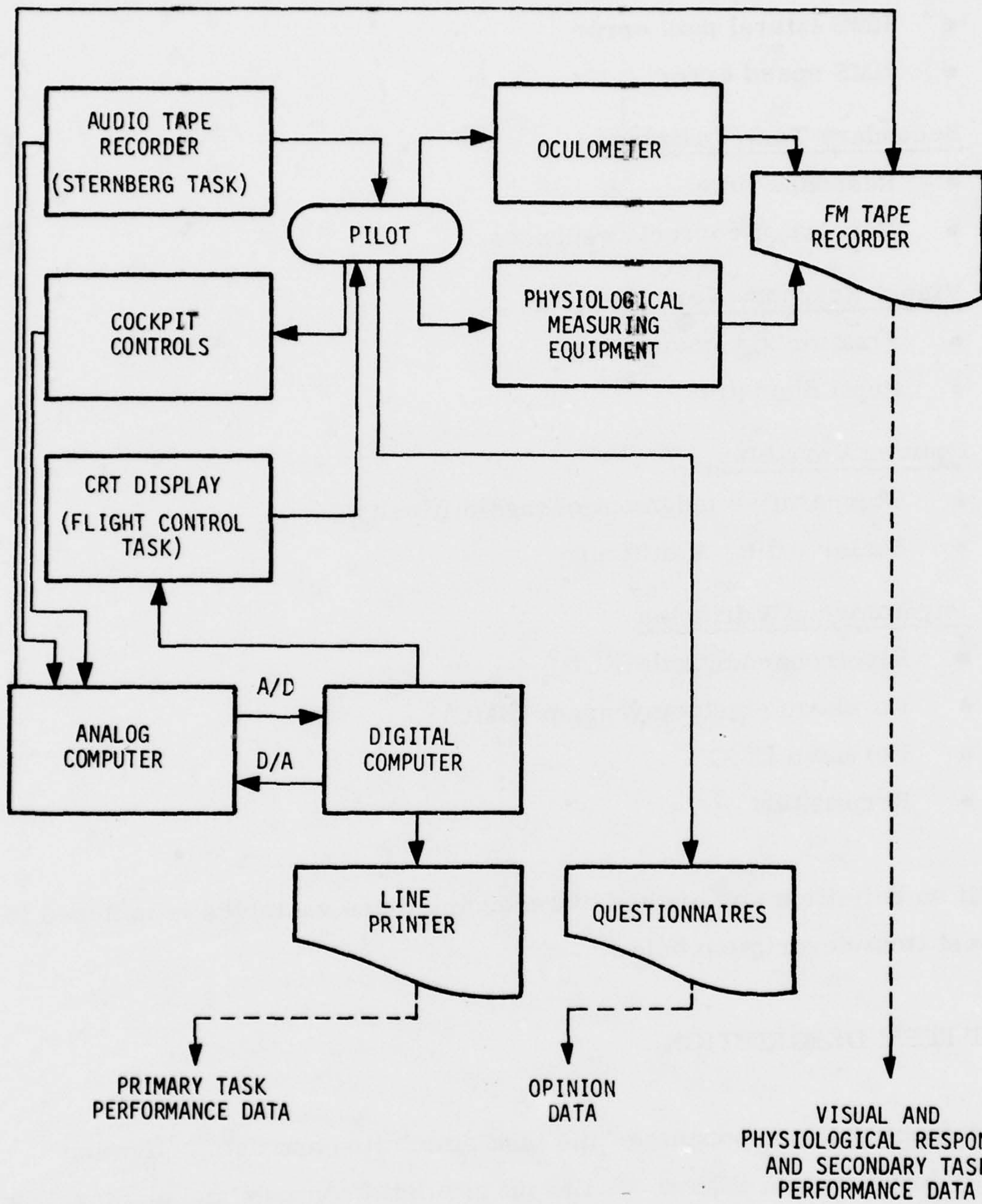


Figure 3. Functional Block Diagram of Simulation Facility

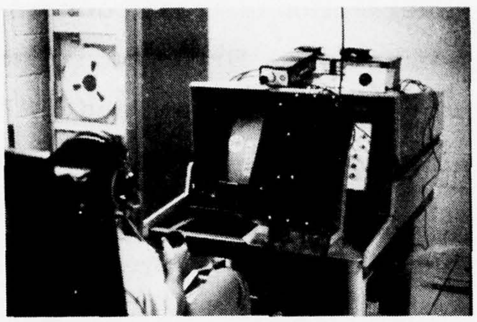


Figure 4. Photographs of Simulator Pilot's Station

Flight control and approach guidance information was presented to the pilot on an electronic vertical situation display (VSD) and rate-of-climb indicator drawn on the CRT. A digital readout of range (feet) to glide path intercept point was also displayed directly above the VSD. Figure 5 illustrates these flight displays. Moving tapes at left, right, and lower edges of the VSD indicate airspeed (knots), altitude (feet x 100), and heading, respectively. Other basic display elements are aircraft symbol, artificial horizon line, pitch and roll attitude scales, and flight path error symbol. The error symbol was driven by lateral and vertical flight-path error signals, and is functionally equivalent to the intersection of cross pointers on conventional horizontal situation or course indicators. Error symbol scaling was 3.85 degrees/inch of display for lateral path deviations and 0.77 degrees/inch for vertical path deviations. These values approximate scale factors used on conventional cross pointers. Computed steering or flight director commands were not displayed to the pilot in this study.

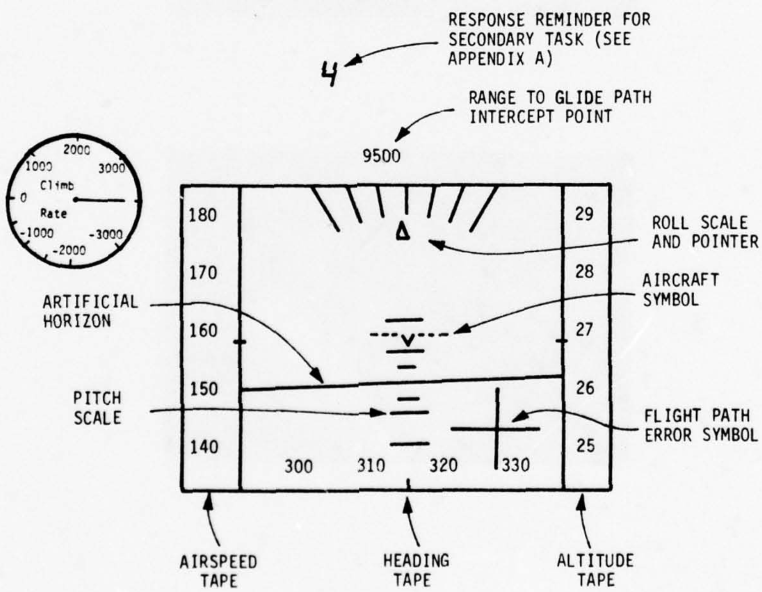


Figure 5. Pilot's Flight Displays

The F-4 aircraft model and flight control system simulated were previously adapted from References 10 and 11. With the exception of pilot's control station and control-gain scaling, the aircraft simulation was all digital. The aircraft was configured with full flaps and landing gear down, and trimmed with initial conditions of level flight at 165 knots in this study. Aerodynamic derivatives accounted for effects of high angle of attack present near stall conditions. The control system included three-axis stability augmentation and turn coordination. Since the pilot's control station had spring-load force-feel characteristics which were not adjustable, control feel characteristics from Reference 11 were not included in the simulated control system. Control surface rate limits were implemented by rate limiting stabilator and aileron commands generated by the control system.

A simplified gust model produced random lateral and vertical perturbations on the aircraft. This model is:

$$\text{gust (ft/sec)} = x\sqrt{\frac{2(1750)}{v}}\left(\frac{1}{(1750/v)(S+1)}\right)$$

where

x = gaussian random number with zero mean, and σ adjusted to produce desired gust amplitudes

v = aircraft velocity, ft/sec

RMS attitudes, path errors, and speed error were computed on each trial. In general,

$$\text{RMS X} = \sqrt{\frac{\overset{N}{\Sigma} X^2}{N}}$$

where

X = performance variable sampled

N = number of samples

Secondary Task Simulation

The Air Force Aerospace Medical Research Laboratory provided copies of tapes containing the secondary task stimuli. Stimulus characteristics are summarized below:

- Positive subset for M = 2: letters A and H
- Positive subset for M = 4: letters A, H, J, and Q
- Negative subset for both of above: letters B, C, E, F, G, I, L, R, and Y
- Mean interstimulus interval: approximately 5.5 seconds
- Range of interstimulus intervals: approximately 2 to 7 seconds
- Random sequence of positive and negative stimuli with constraint that probability of positive stimulus is P = 0.5

Stimuli were presented through headsets worn by the simulator pilot. The pilot was instructed to respond by appropriate activation of the pitch-trim switch--forward for positive and aft for negative stimuli.

The interval between stimulus presentation and either a correct or incorrect response defined response time on this task. Correctness of response was scored primarily to verify that subjects were responding to the secondary

task as instructed. Since percent of correct responses was expected to be near or at the 100 percent level, this measure was not anticipated to be a useful task performance measure.

Visual Response Data Recording

A Honeywell Mark 3A remote oculometer system calculated the pilot's eye fixations and pupil diameter during each test trial. The system's electro-optical unit, which illuminates the eye and images the illuminated eye onto an IR vidicon, was mounted adjacent to the 19-inch CRT primary task display. Location of this unit is shown in Figure 4.

Opinion Questionnaires

Two questionnaires were administered to obtain pilots' subjective ratings of task difficulty or workload. Paired-comparison judgments of relative task difficulty were requested after successive pairs of trials. Pilots were asked simply to "indicate which of the two preceding trials imposed the highest overall workload level."

The second questionnaire (see Figure 6) is a form administered twice to each pilot for each of the six test conditions to obtain a scalar rating of overall task workload. This form is based on the Cooper-Harper rating scale for handling qualities (Reference 12) which was modified for purposes of the present study to focus on task workload rather than aircraft handling qualities. It must be emphasized that the form as modified in Figure 6 has not been validated as a workload rating scale. Inferences of this modified form's validity should not be made based on extensive previous work and experience with the Cooper-Harper scale.

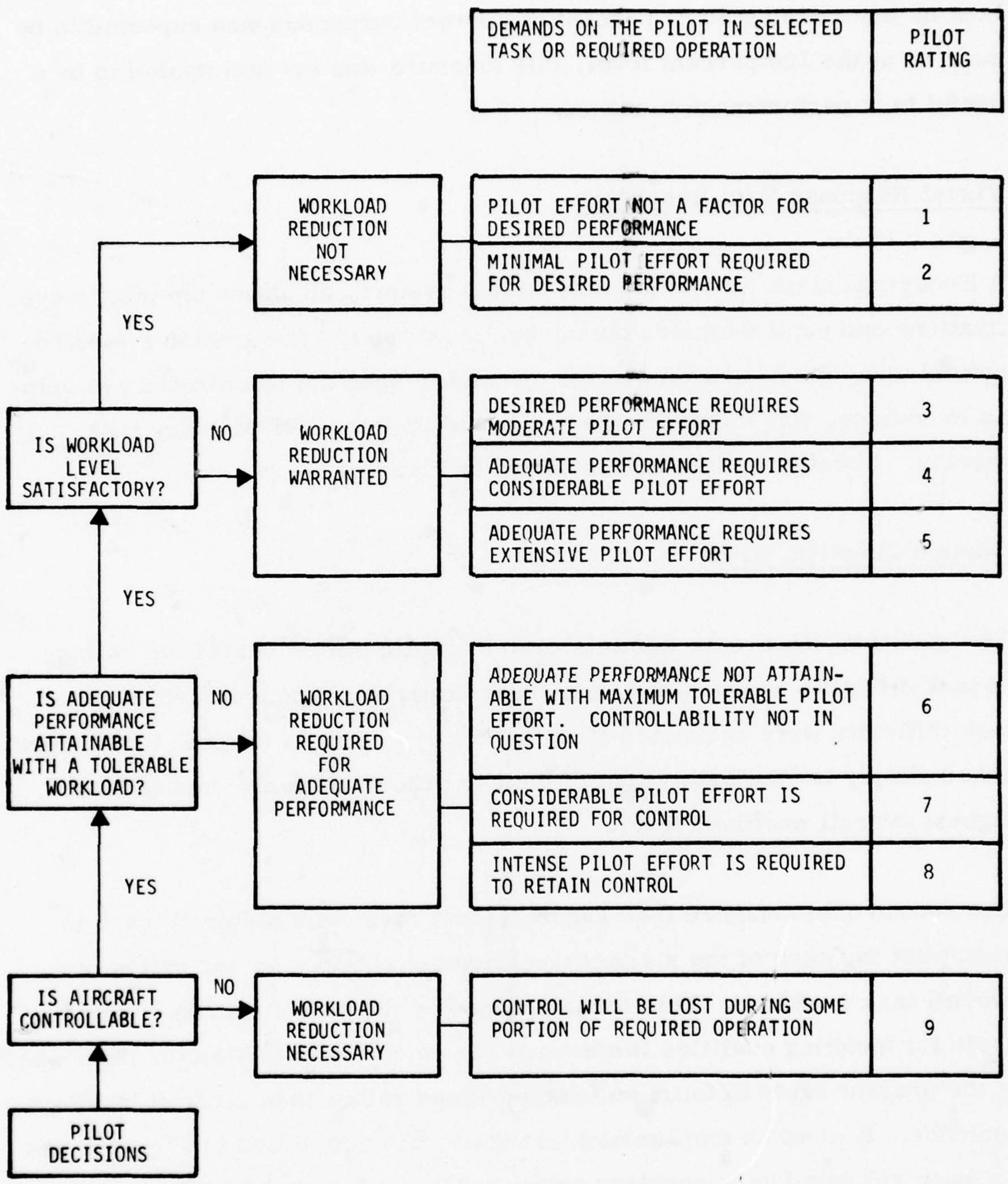


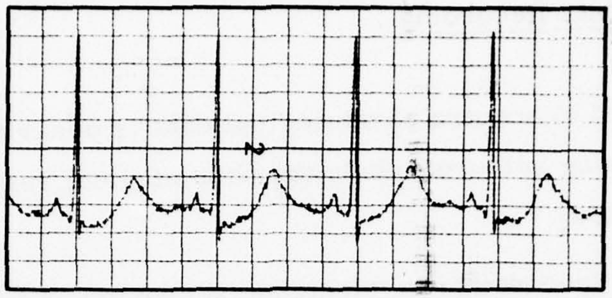
Figure 6. Workload Rating Scale (Adapted from Reference 12)

Physiological Data Recording

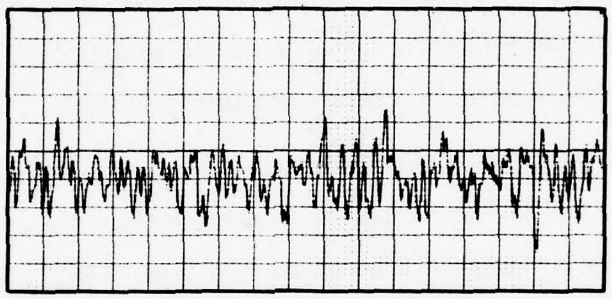
The system used to acquire physiological data is described in Reference 2. Methods of sensing and processing the physiological response signals are summarized below:

- ECG--Recorded from electrode approximately three inches below front edge of right armpit and reference electrode at waist level on subject's right side. Electrode outputs processed by Honeywell Accudata 135A biomedical amplifier with ECG isolator. Low and high frequency filters set at 0.05 Hz and 100 Hz, respectively. Overall voltage gain of approximately 2000:1.
- EMG--Recorded from electrodes on forehead just above eyebrows and on right forearm (two data channels). Electrode outputs processed by Accudata 135A amplifier with EEG/EMG preamplifier. Low and high frequency filters set at 50 Hz and 2500 Hz, respectively. Overall voltage gain of approximately 1000:1.
- Respiration--Recorded from mercury chestband respiration transducer, and amplified with Accudata 137 respiration control/tachometer. Waveform output was used.

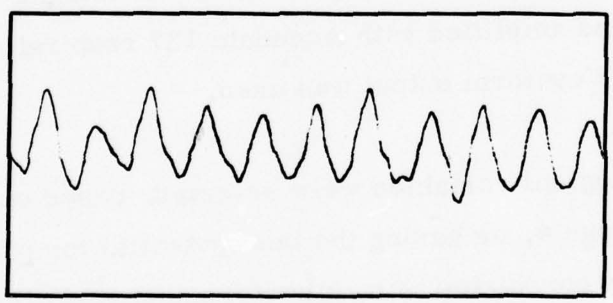
The above physiological variables were selected, based on previous work in References 1 through 4, as having the best potential for producing extracted features which are correlates of task performance and pilot-opinion of task difficulty. Sample time histories of ECG, EMG, and respiration signals are shown in Figure 7.



a) ECG Sample Time History (25 mm/sec)



b) Arm EMG Sample Time History (125 mm/sec)



c) Respiration Sample Time History (2.5 mm/sec)

Figure 7. ECG, EMG, and Respiration Sample Time Histories

TEST PROCEDURES

Schedule

Each subject was scheduled for a three-day period planned as indicated below.

Day 1

1300 to 1400:	Briefing
1400 to 1700:	Informal practice, begin formal practice

Day 2

0900 to 1200:	Conclude formal practice and begin data collection
1400 to 1700:	Conclude data collection

Day 3

0900 to 1200:	Contingency time in event of schedule slippage

Briefing

Written instructions summarizing task procedures (see Appendix A) were included as part of a briefing package given to subjects for review. Instructions necessarily identified the two secondary task conditions, but identified only the variables (turbulence and FCS characteristics) involved in creating a range of primary-task difficulty. Thus the subjects were never informed as to the actual number of different test conditions.

Informal Practice

Approximately two hours were devoted to informal practice and familiarization with primary and secondary tasks.

Formal Practice

Formal practice followed the same procedures to be applied during data collection, including adherence to prescribed separately-randomized pairs of test conditions and administration of opninon questionnaires. Each subject received four formal-practice trials per test condition, or a total of 24 trials.

Data Collection

Data collection was conducted according to the following procedural sequence:

1. Check calibration and operation of all recording equipment
2. Complete 12 data trials (nominally two minutes each), allowing 30-second inter-trial intervals
3. Minimum 20 minute rest interval
4. Twelve data trials
5. Twenty-minute rest
6. Twelve data trials
7. Twenty-minute rest

8. Twelve data trials

9. Twenty-minute rest

10. Twelve data trials

Prior to each trial, the pilot was told which secondary-task response would be required. Paired-comparison judgments on overall task workload were requested after each pair of trials. Scalar ratings were requested once after each test condition in block 4 above, and again once for each condition in block 10.

DATA ANALYSIS PROCEDURES

The statistical analysis procedures outlined in Figure 8 are similar in most respects to those applied in References 2, 3, and 4. Analog data tapes are sampled and selected features are extracted. Physiological features are simple statistical summaries (i.e., mean and standard deviations) of selected waveform characteristics computed over some interval. In the present study, this interval was the final 90 seconds of a test trial.

Feature data obtained from analog tape-recorded signals and other response data input directly from punch cards are then normalized by

$$Z_{ijk} = \frac{X_{ijk} - \overline{X}_{jk}}{\sigma_{x_{jk}}}$$

where

X_{ijk} = raw score value i on response measure j for subject k

Z_{ijk} = normalized value of X_{ijk}

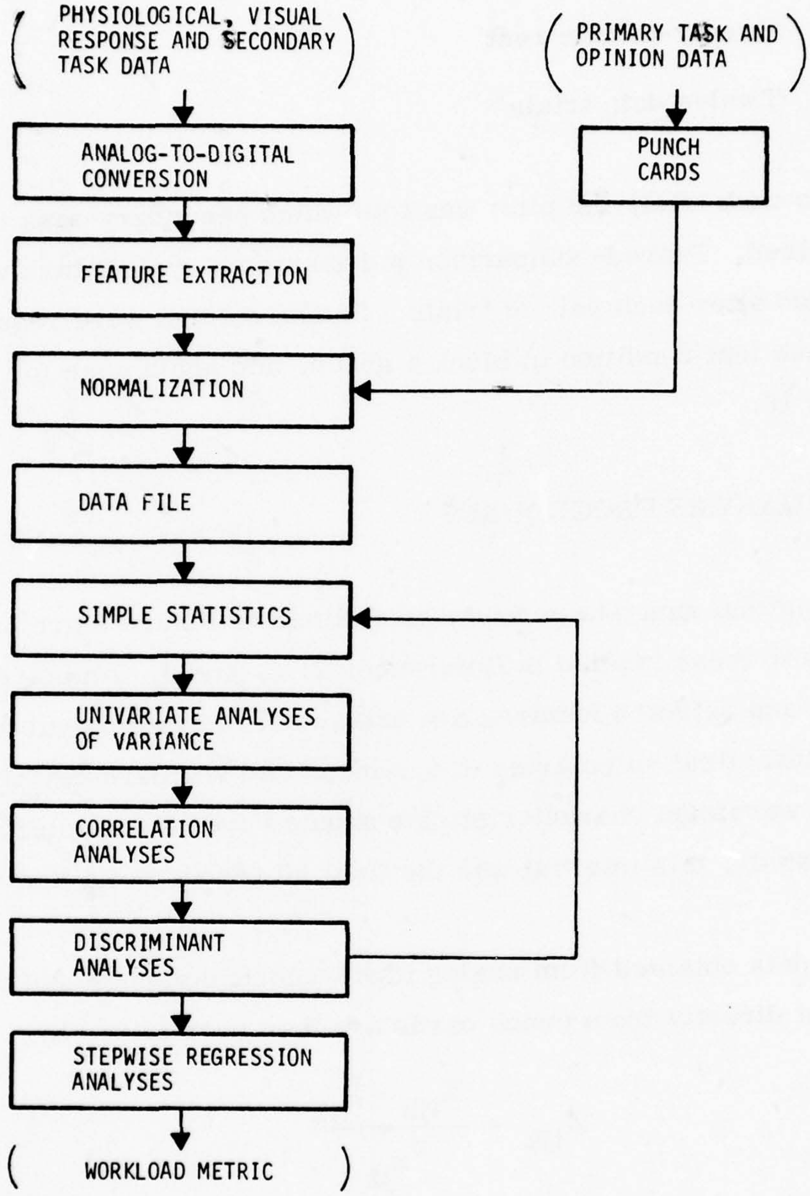


Figure 8. Outline of Analysis Procedure

$\overline{X}_{jk}$ = mean raw score value on measure j for subject k

$\sigma_{x_{jk}}$ = standard deviation of raw scores for measure j and subject k

Normalization in the above manner generates a transformed set of data which places all response measures on a common scale with zero mean and unity standard deviation for each measure/subject combination. The resulting file containing all normalized data provides the data base for all subsequent analyses. Simple statistics (mean and standard deviations), univariate analyses of variance, and correlation coefficients are computed on all variables in the data base to summarize effects of test conditions and interrelationships between variables.

Discriminant and regression techniques are the primary statistical tools applied in the analysis procedure. These techniques are summarized below and described in greater detail in References 13, 14, and 15.

Discriminant Analysis

Two functionally different sets of dependent variables were measured in the present study. One set includes the physiological and visual response variables representing the potential source of a workload metric. The other set includes performance and opinion measures recorded on tasks for which there was a priori evidence of differences in workload. A major goal was to derive weightings on the physiological and visual response variables that reflect maximum differences between the workload levels imposed by the six test conditions.

The first step in this process is to derive linear combinations of performance and opinion scores that best differentiate between the six conditions. This is accomplished by multiple discriminant analysis. A discriminant function resulting from this analysis defines a single scale composed of the original set of performance and opinion data, and can be expressed by the general form

$$D_i = \sum^{n} (a_j Z_{ij}) \tag{1}$$

where

D_i = score i on discriminant scale D

Z_{ij} = normalized score Z_i on performance or opinion variable j

a_j = weighting coefficient on variable j

n = number of variables included in the discriminant analysis

In effect, values of D for equation (1) create a "new" variable which is a composite based on the performance and opinion variables entered into the analysis. Advantages of this technique are that the information content in each variable is used toward maximization of group (condition) differences, and the relative contribution of each variable can be assessed. Composite variables generated in this manner were statistically analyzed in the same way as the individual variables constituting the discriminant function (see Figure 8).

Discriminant analyses were performed on the following variable combinations to allow assessment of possible differences in regression-analysis results due to the discriminant function selected.

- Primary task measures only
- Opinion measures only
- Primary and secondary task measures combined
- Primary task, secondary task, and opinion measures combined

Stepwise Regression Analysis

A discriminant function produced by the process described above can be interpreted as a scale of difficulty or workload imposed by the various experimental conditions. A discriminant score derived from performance and opinion data on a particular trial represents a position on this scale relative to scores from all other trials. Using the discriminant score as a criterion measure, stepwise multiple regression may be used to derive the predictive relationship between the physiological/visual response measures and this criterion score.

Stepwise regression is a procedure for selective examination of the available predictor variables for their individual contributions to the explanation of variance in the criterion measure. The variables that are included in the final equation are the set with the highest unique contribution to prediction of criterion variability. At each step, a partial F test for each variable is computed and compared to a preselected value. A variable not yet in the equation which provides the greatest significant contribution is added

to the equation. As variables are added, a variable placed in the equation on previous steps may no longer provide significant unique contribution to prediction of criterion variability. If so, that variable is removed. This stepwise process is continued until none of the predictor variables can be incorporated into, or removed from, the regression equation based on the partial F test.

The result is a linear combination of a minimum set of predictor variables and associated weighting coefficients which best predict response on the criterion variable. Expressed as a prediction equation, this result is

$$\hat{D}_i = \sum^{m} (b_j Z_{ij}) \tag{2}$$

where

$\hat{D}_i$ = predicted score i on discriminant scale D

Z_{ij} = normalized score Z_i on physiological or visual response variable j

b_j = weighting coefficient on variable j

m = number of variables providing significant unique contribution to prediction of scores on discriminant scale D

Within the conceptual framework of this study, equation (2) above operationally defines the form of a workload metric with at least ordinal characteristics. Thus,

$$W = \sum^{m} (b_j Z_{ij}) \tag{3}$$

where

W = workload metric

Regression analyses were performed on several combinations of predictor variables to identify a combination which best characterized workload with as few variables from as few sources as possible. Variable combinations analyzed are discussed in Section III.

VARIABLES ANALYZED

Variables included in the above statistical analyses are listed and described in Table 1. The following examples aid in understanding the mnemonics applied:

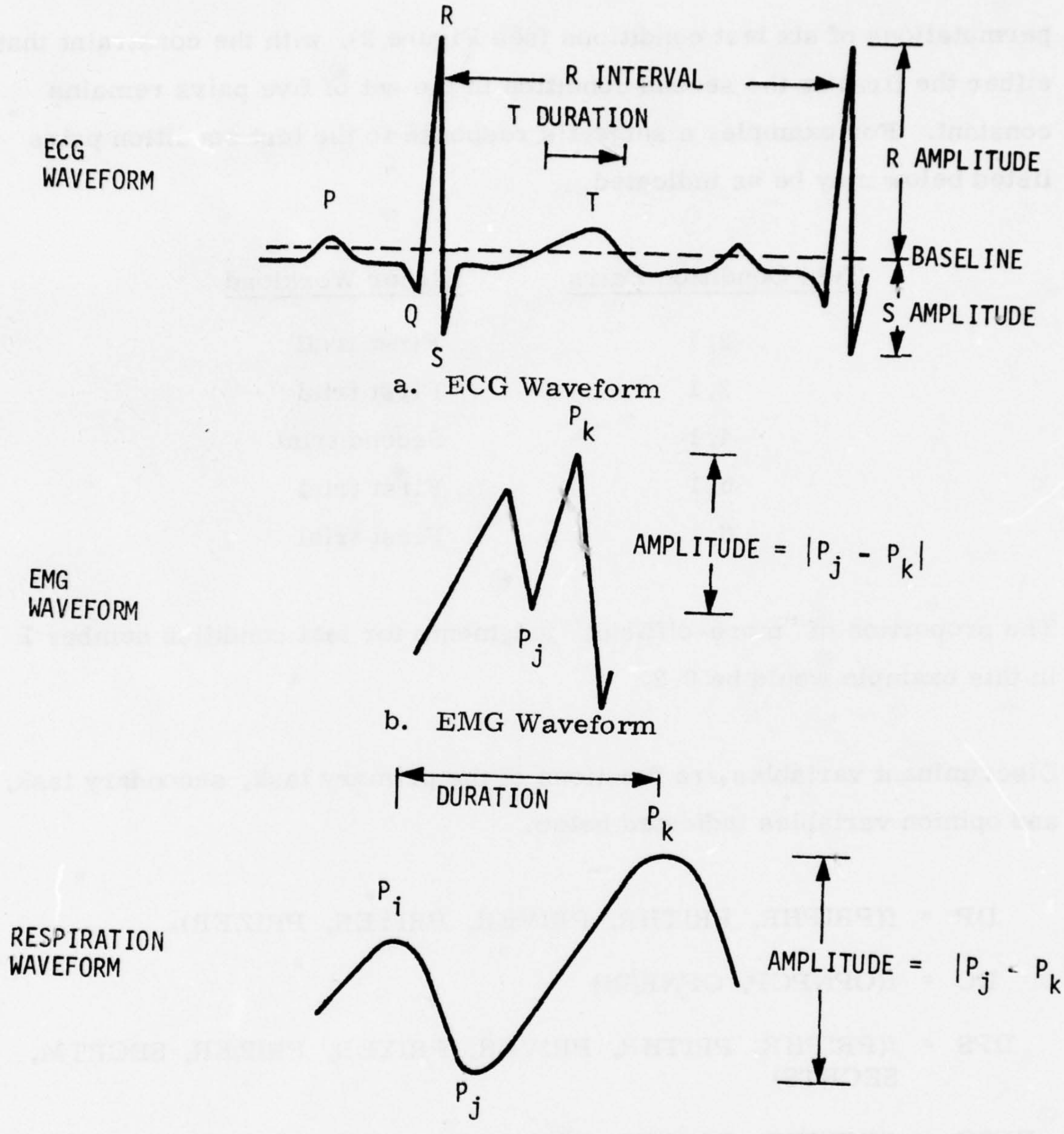
- EMGHAM--EMG, head, amplitude, mean
- PRITHR--primary task, pitch attitude (theta), RMS
- OPNRSN--opinion, rating scale, numeric rating

Physiological features represented in variables 1 through 19 are clarified in Figure 9. ECG waveform amplitudes are defined relative to a common baseline which is the mean signal level recorded on each trial (see Figure 9 examples for R- and S-wave amplitudes). Samples of EMG and respiration amplitude are defined by absolute value of the difference between consecutive peaks (slope reversals). The ECG R-wave interval is the duration between R-wave peaks in this periodic waveform. Respiration duration is defined in a similar manner as the interval between successive signal peaks in the same direction.

Variables 20 and 21 are summary indicators of eye motion activity based on a velocity vector computed from x, y coordinate time histories. Each value of variable 30 is based on five comparative judgments representing possible

TABLE 1. VARIABLES ANALYZED

Variable Number	Mnemonic	Description
1	EMGHAM	Mean EMG amplitude; head
2	EMGHAS	Standard deviation EMG amplitude; head
3	EMGAAM	Mean EMG amplitude; arm
4	EMGAAS	Standard deviation EMG amplitude; arm
5	RESPAM	Mean respiration amplitude
6	RESPAS	Standard deviation respiration amplitude
7	RESPDM	Mean respiration duration
8	RESPDS	Standard deviation respiration duration
9	ECGRAM	Mean ECG R-wave amplitude
10	ECGRAS	Standard deviation ECG R-wave amplitude
11	ECGRIM	Mean ECG R-wave interval
12	ECGRIS	Standard deviation ECG R-wave interval
13	ECGQRAM	Mean ECG Q/R-wave amplitude ratio
14	ECGSRAM	Mean ECG S/R-wave amplitude ratio
15	ECGTRAM	Mean ECG T/R-wave amplitude ratio
16	ECGQDM	Mean ECG Q-wave duration
17	ECGRDM	Mean ECG R-wave duration
18	ECGSDM	Mean ECG S-wave duration
19	ECGTDM	Mean ECG T-wave duration
20	EYESVM	Mean eye scan velocity
21	EYESVS	Standard deviation eye scan velocity
22	EYEPDM	Mean eye pupil diameter
23	SECRTM	Mean secondary task response time
24	SECRTS	Standard deviation secondary task response time
25	PRIPHR	RMS primary task roll attitude
26	PRITHR	RMS primary task pitch attitude
27	PRIVER	RMS primary task speed error
28	PRIYER	RMS primary task lateral path error
29	PRIZER	RMS primary task vertical path error
30	OPNPCP	Proportion more-difficult judgments; paired comparison opinion
31	OPNRSN	Numeric rating; rating scale opinion
32	DP	Discriminant scale; primary task measures only
33	DO	Discriminant scale; opinion measures only
34	DPS	Discriminant scale; primary and secondary task measures combined
35	DPSO	Discriminant scale; primary task, secondary task, and opinion measures combined



a. ECG Waveform

b. EMG Waveform

c. Respiration Waveform

Figure 9. ECG, EMG, and Respiration Waveform Features

permutations of six test conditions (see Figure 2), with the constraint that either the first or the second condition in the set of five pairs remains constant. For example, a subject's response to the test condition pairs listed below may be as indicated.

Test Condition Pairs	Higher Workload
2, 1	First trial
3, 1	First trial
4, 1	Second trial
5, 1	First trial
6, 1	First trial

The proportion of "more-difficult" judgments for test condition number 1 in this example would be 0.2.

Discriminant variables are functions of the primary task, secondary task, and opinion variables indicated below.

DP = f(PRIPHR, PRITHR, PRIVER, PRIYER, PRIZER)

DO = f(OPNPCP, OPNRSN)

DPS = f(PRIPHR, PRITHR, PRIVER, PRIYER, PRIZER, SECRTM, SECRTS)

DPSO = f(PRIPHR, PRITHR, PRIVER, PRIYER, PRIZER, SECRTM, SECRTS, OPNPCP, OPNRSN)

SECTION III

RESULTS

EFFECTS OF TEST CONDITIONS

Table 2 summarizes results of the univariate analyses of variance performed on all variables. Significant results at the level of $p < 0.05$ are indicated for the primary-task main effect (P), the secondary-task main effect (S), and the P x S interaction. Appendix B contains plots of normalized data for all variables showing statistically significant effects due to the test conditions.

Primary-task conditions produced significant changes in a number of physiological, visual, performance, and opinion responses (see Table 2). Appendix B indicates these effects to be most substantial for EMG (arm), respiration, pupil-diameter, primary-task, and opinion measures. The direction of these effects is as anticipated.

Generally, the secondary task conditions had only minor and inconsistent effects in comparison to effects of the primary task. Main effects for the secondary task which are significant are limited to two of the composite variables generated from discriminant analysis. These differences, however, are not in the anticipated direction (e.g., see Figure B-23). Lack of consistency in the apparent effects of S is reflected in the P x S interactions.

TABLE 2. UNIVARIATE ANALYSIS OF VARIANCE RESULTS

	Effects of Test Conditions			Appendix B Figure No.
	P	S	PxS	
1. EMGHAM				
2. EMGHAS				
3. EMGAAM	*			B1
4. EMGAAS	*			B2
5. RESPAM				
6. RESPAS	*			B3
7. RESPDM	*		*	B4
8. RESPDS	*			B5
9. ECGRAM			*	B6
10. ECGRAS	*		*	B7
11. ECGRIM				
12. ECGRIS	*			B8
13. ECGQRAM				
14. ECGSRAM				
15. ECGTRAM				
16. ECGQDM			*	B9
17. ECGRDM				
18. ECGSDM				
19. ECGTDM				
20. EYESVM				
21. EYESVS				
22. EYEPDM	*			B10
23. SECRTM	*			B11
24. SECRTS	*			B12
25. PRIPHR	*			B13
26. PRITHR	*			B14
27. PRIVER	*		*	B15
28. PRIYER	*			B16
29. PRIZER	*		*	B17
30. OPNPCP	*			B18
31. OPNRSN	*			B19
32. DP	*			B20
33. DO	*	*	*	B21
34. DPS	*			B22
35. DPSO	*	*	*	B23

*Significant at $p < 0.05$

These findings suggest that differences in overall task difficulty due to secondary task conditions (M = 2 and M - 4) were relatively small compared to the trial-to-trial variations in flight-task difficulty experienced by the subjects under a single primary task condition.

CORRELATIONS BETWEEN VARIABLES

Correlations (r) between all variables analyzed are listed in Table 3. The following observations from this table aid interpretation of other analysis results.

Correlations between variables showing similar trends in Appendix B would be expected. For example, in comparison to some other physiological variables, arm EMG variables (numbers 3 and 4) show relatively high correlations with primary task and opinion variables. Arm EMG response would therefore be anticipated to receive relatively high loadings in regression results. However, the two arm EMG variables are themselves highly correlated ($r = 0.96$). Since the stepwise regression procedure maintains only those predictors accounting for significant unique variance in the criterion, only one of the arm EMG variables would be expected to "survive" the stepwise process.

Secondary task, primary task, and opinion variables (numbers 23 through 31) were applied to derive discriminant functions (variables 32 through 35). Correlations of secondary task and discriminant variables are relatively low, indicating that secondary task variables should have been assigned relatively low weightings in the discriminant functions.

TABLE 3. CORRELATIONS BETWEEN ALL VARIABLES

Variables	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 EMGHAM	0.71	-0.11	-0.10	-0.05	0.02	0.12	0.12	-0.06	-0.11	-0.03	0.01	-0.12	0.02	0.04	0.01	-0.08	-0.03
2 EMGHAS		-0.10	-0.09	0.03	0.14	0.12	0.16	-0.01	-0.03	-0.05	0.07	-0.05	0.00	-0.08	0.02	-0.05	-0.00
3 EMGAAM			0.96	0.06	0.14	-0.38	-0.02	-0.04	0.16	-0.08	0.21	0.10	0.07	-0.05	-0.13	0.05	0.11
4 EMGAAS				0.06	0.18	-0.38	0.03	-0.06	0.15	-0.08	0.20	0.11	0.04	-0.08	-0.13	0.04	0.10
5 RESPAM					0.23	0.21	0.04	0.26	0.15	0.16	0.13	-0.09	0.20	0.02	0.12	0.12	0.13
6 RESPAS						-0.03	0.54	0.11	0.15	0.03	0.15	-0.07	0.08	0.04	0.03	0.06	0.01
7 RESPDM							0.33	0.09	-0.09	0.18	-0.08	-0.12	-0.03	0.10	0.19	-0.14	-0.03
8 RESPDS								0.07	0.08	0.04	0.12	-0.04	0.00	0.06	0.04	0.02	0.01
9 ECGRAM									-0.04	0.33	-0.09	-0.02	0.34	0.23	0.36	0.18	0.32
10 ECGRAS										0.01	0.21	0.04	0.10	0.01	-0.00	0.10	-0.09
11 ECGRIM											0.10	-0.22	0.33	0.39	0.40	0.18	0.31
12 ECGRIS												-0.07	-0.07	-0.18	-0.15	-0.03	0.13
13 ECGQRAM													-0.08	-0.44	-0.36	0.29	0.04
14 ECGSRAM														0.17	0.27	0.47	0.13
15 ECGTRAM															0.45	-0.12	-0.09
16 ECGQDM																0.04	0.04
17 ECGRDM																	0.23
18 ECGSDM																	
19 ECGTDM																	
20 EYESVM																	
21 EYESVS																	
22 EYEPDM																	
23 SECRTM																	
24 SECRTS																	
25 PRIPHR																	
26 PRITHR																	
27 PRIVER																	
28 PRIYER																	
29 PRIZER																	
30 OPNPCP																	
31 OPNRSN																	
32 DP																	
33 DO																	
34 DPS																	
35 DPSO																	

TABLE 3. CORRELATIONS BETWEEN ALL VARIABLES (concluded)

Variables	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1 EMGHAM	0.04	0.04	0.04	0.02	-0.01	-0.06	-0.04	-0.01	0.03	0.03	0.03	-0.01	-0.01	-0.02	-0.01	-0.03	-0.02
2 EMGHAS	0.03	0.02	0.02	-0.04	-0.06	-0.05	-0.06	-0.09	-0.03	0.03	-0.01	-0.04	-0.07	-0.07	-0.04	-0.07	-0.06
3 EMGAAM	-0.15	-0.08	-0.12	0.36	0.16	0.23	0.56	0.46	0.40	0.34	0.34	0.50	0.43	0.57	0.51	0.57	0.55
4 EMGAAS	-0.12	-0.10	-0.14	0.33	0.20	0.26	0.56	0.45	0.40	0.36	0.31	0.48	0.41	0.56	0.49	0.56	0.53
5 RESPAM	-0.03	-0.03	-0.04	0.15	0.10	0.14	0.10	0.00	0.05	0.09	0.06	0.07	0.06	0.08	0.07	0.08	0.08
6 RESPAS	0.02	0.10	0.02	0.08	0.15	0.15	0.30	0.18	0.21	0.25	0.16	0.21	0.24	0.28	0.23	0.28	0.26
7 RESPDM	0.14	0.08	0.09	-0.35	-0.10	-0.12	-0.37	-0.28	-0.26	-0.26	-0.22	-0.35	-0.33	-0.36	-0.36	-0.36	-0.38
8 RESPDS	0.01	0.13	0.06	-0.05	0.06	0.02	0.11	0.03	0.10	0.09	0.05	0.06	0.08	0.10	0.07	0.10	0.08
9 ECGRAM	0.05	0.05	0.01	-0.05	-0.00	0.00	-0.01	-0.06	-0.10	-0.04	-0.07	-0.01	-0.05	-0.03	-0.02	-0.03	-0.02
10 ECGRAS	0.05	0.01	0.02	0.13	0.07	0.07	0.19	0.13	0.16	0.20	0.12	0.14	0.11	0.18	0.14	0.18	0.16
11 ECGRIM	0.21	-0.13	-0.15	-0.01	-0.02	0.01	-0.07	0.01	-0.03	-0.09	-0.12	-0.05	-0.03	-0.06	-0.05	-0.06	-0.06
12 ECGRIS	-0.17	-0.06	0.03	0.11	0.12	0.10	0.13	0.14	0.10	0.12	0.17	0.09	0.13	0.15	0.10	0.15	0.12
13 ECGQRAM	-0.26	-0.16	-0.11	-0.03	-0.01	-0.01	-0.07	-0.05	-0.02	-0.08	0.03	-0.02	-0.06	-0.05	-0.03	-0.05	-0.04
14 ECGSRAM	0.02	-0.07	-0.13	0.09	-0.05	-0.00	-0.04	-0.02	-0.06	-0.10	-0.08	-0.04	0.00	-0.04	-0.04	-0.04	-0.04
15 ECGTRAM	0.38	0.19	0.05	0.12	-0.16	-0.06	-0.02	-0.05	-0.04	-0.04	-0.09	-0.03	-0.02	-0.04	-0.03	-0.04	-0.04
16 ECGQDM	0.28	0.08	0.04	-0.06	-0.08	-0.03	-0.06	-0.08	-0.08	-0.06	-0.10	-0.06	-0.03	-0.08	-0.06	-0.08	-0.07
17 ECGRDM	-0.12	-0.12	-0.09	0.09	0.06	0.04	-0.01	-0.03	-0.02	-0.05	-0.00	0.01	0.00	-0.01	0.01	-0.01	0.01
18 ECGSDM	-0.51	-0.15	-0.12	0.05	0.06	0.06	0.04	0.08	0.01	-0.01	-0.01	0.01	0.01	0.04	0.01	0.04	0.02
19 ECGTDM		0.06	0.01	0.05	-0.09	-0.04	-0.03	-0.01	-0.01	0.05	-0.03	-0.02	-0.03	-0.04	-0.02	-0.04	-0.03
20 EYESVM			0.85	-0.06	0.04	0.09	0.06	-0.04	0.10	0.03	0.13	0.05	0.03	0.07	0.05	0.08	0.06
21 EYESVS				-0.17	0.04	0.07	-0.00	-0.07	0.04	0.03	0.09	-0.05	-0.06	0.01	-0.05	0.01	-0.04
22 EYEPDM					0.10	0.16	0.31	0.24	0.21	0.19	0.23	0.29	0.27	0.32	0.30	0.32	0.31
23 SECRTM						0.67	0.32	0.23	0.28	0.29	0.25	0.26	0.23	0.32	0.27	0.32	0.30
24 SECRTS							0.31	0.25	0.32	0.29	0.26	0.28	0.24	0.32	0.29	0.34	0.31
25 PRIPHR								0.64	0.64	0.64	0.55	0.79	0.65	0.97	0.80	0.97	0.89
26 PRITHR									0.60	0.41	0.38	0.61	0.54	0.72	0.62	0.72	0.68
27 PRIVER										0.47	0.62	0.63	0.57	0.75	0.65	0.74	0.70
28 PRIYER											0.40	0.47	0.41	0.57	0.48	0.57	0.53
29 PRIZER												0.59	0.49	0.68	0.60	0.68	0.64
30 OPNPCP													0.71	0.82	0.99	0.82	0.97
31 OPNRSN														0.69	0.82	0.69	0.80
32 DP															0.84	1.00	0.92
33 DO																0.84	0.98
34 DPS																	0.92
35 DPSO																	

Of particular interest in the correlations between discriminant variables is the value of $r = 0.84$ between variables 32 (discriminant on primary task measures only) and 33 (discriminant on opinion measures only). Discriminant scores produced by the two independent sets of measures are closely correlated.

DISCRIMINANT FUNCTIONS

Coefficients obtained from the four discriminant analyses performed on different combinations of variables are listed in Table 4. Each analysis produced a number of uncorrelated functions (see Reference 15, p. 162). The table includes coefficients for only the first discriminant function yielded by each analysis. The percent variance indicated in Table 4 is the portion of variance accounted for by the first discriminant function relative to the total variance accounted for by all functions generated in each analysis.

Secondary task variables have low weightings in comparison to primary task and opinion measures. Variables consistently having the highest weightings are numbers 25 (RMS roll attitude) and 30 (paired-comparison opinion of task difficulty). Thus, where included in an analysis, these response variables provide the greatest individual contributions to defining scores on the discriminant or composite variables.

TABLE 4. DISCRIMINANT FUNCTION COEFFICIENTS

Variables	Discriminant Functions			
	DP	DO	DPS	DPSO
23. SECRTM			-0.024	0.010
24. SECRTS			0.034	0.001
25. PRIPHR	0.949		0.949	0.426
26. PRITHR	0.157		0.156	0.053
27. PRIVER	0.123		0.119	0.048
28. PRIYER	-0.141		-0.142	-0.070
29. PRIZER	0.199		0.198	0.050
30. OPNPCP		0.964		0.868
31. OPNRSN		0.266		0.232
(% variance)	(97.6)	(99.9)	(97.1)	(98.4)

REGRESSION ANALYSIS RESULTS

As previously noted in Section II, regression analyses were performed on several combinations of predictor variables to identify a combination which best characterized workload with a minimal number of variables and measurement sources.

Stepwise regression coefficients were initially computed for a baseline condition utilizing all physiological and visual response variables as predictors and the discriminant variables described above as criterion variables. Resulting coefficients are listed in the first column of Tables 5 through 8. Zeros are shown to indicate variables included in the stepwise process, but identified as not making a unique significant contribution to

prediction of variance in the criterion measure. The coefficient of determination, R^2, indicates that proportion of variance in the criterion accounted for by a weighted linear combination of the predictors. Weights are the coefficients listed in Tables 5 through 8.

Analysis on the complete set of predictors yielded generally higher weightings on physiological variables, and in particular, arm EMG and respiration measures. Based on this finding, additional analyses were performed on the set of all physiological variables and a selected subset of these consisting of only arm EMG and respiration variables. Resulting coefficients are listed in the second and third columns of Tables 5 through 8.

Influences on R^2 due to the substantial reduction in the number of predictors are minimal. Comparing R^2 for the complete set versus the smallest selected subset of predictors, reduction in variance accounted for averages less than 4 percent. This finding is encouraging. Measurement sources needed to provide data for the smallest group of predictors include only two electrodes (for arm EMG) and a mercury strain gauge (for respiration).

Weighting coefficients follow a similar pattern on the selected group of predictor variables. The discriminant variable, DPSO, is taken as the most representative criterion since it includes both performance and opinion measures sensitive to task difficulty.

The following expression defines the predicted value of DPSO from coefficients in Table 8:

$$\begin{aligned} \text{DPSO} = {} & 0.631\,(\text{EMGAAM}) + 0.103\,(\text{RESPAM}) + 0.163\,(\text{RESPAS}) \\ & - 0.386\,(\text{RESPDM}) + 0.167\,(\text{RESPDS}) \end{aligned}$$

TABLE 5. REGRESSION COEFFICIENTS FOR PREDICTION OF DP

	Variables Included in Regression		
	Physiological and Visual Response Variables	Physiological Variables Only	Selected Physiological Variables
1. EMGHAM	0	0	
2. EMGHAS	0	0	
3. EMGAAM	0.535	0.545	0.537
4. EMGAAS	0	0	0
5. RESPAM	0	0.085	0.075
6. RESPAS	0.170	0.149	0.160
7. RESPDM	-0.234	-0.295	-0.282
8. RESPDS	0.099	0.134	0.135
9. ECGRAM	0	0	
10. ECGRAS	0.078	0.075	
11. ECGRIM	0	0	
12. ECGRIS	0	0	
13. ECGQRAM	-0.153	-0.139	
14. ECGSRAM	-0.113	-0.142	
15. ECGTRAM	-0.091	0	
16. ECGQDM	0	0	
17. ECGRDM	0	0	
18. ECGSDM	0	0	
19. ECGTDM	0	0	
20. EYESVM	0.113		
21. EYESVS	0		
22. EYEPDM	0.107		
(R^2)	(0.438)	(0.426)	(0.399)

TABLE 6. REGRESSION COEFFICIENTS FOR PREDICTION OF DO

Predictor Variables	Variables Included in Regression		
	Physiological and Visual Response Variables	Physiological Variables Only	Selected Physiological Variables
1. EMGHAM	0	0	
2. EMGHAS	0	0	
3. EMGAAM	0.467	0.479	0.460
4. EMGAAS	0	0	0
5. RESPAM	0.101	0.104	0.079
6. RESPAS	0.093	0.096	0.109
7. RESPDM	-0.313	-0.327	-0.301
8. RESPDS	0.109	0.135	0.123
9. ECGRAM	0	0	
10. ECGRAS	0	0	
11. ECGRIM	0	0	
12. ECGRIS	0	0	
13. ECGQRAM	-0.093	-0.092	
14. ECGSRAM	-0.129	-0.119	
15. ECGTRAM	0	0	
16. ECGQDM	0	0	
17. ECGRDM	0	0	
18. ECGSDM	0	0	
19. ECGTDM	0	0.074	
20. EYESVM	0.287		
21. EYESVS	-0.251		
22. EYEPDM	0		
(R^2)	(0.358)	(0.345)	(0.325)

TABLE 7. REGRESSION COEFFICIENTS FOR PREDICTION OF DPS

Predictor Variables	Variables Included in Regression		
	Physiological and Visual Response Variables	Physiological Variables Only	Selected Physiological Variables
1. EMGHAM	0	0	
2. EMGHAS	0	0	
3. EMGAAM	0.536	0.546	0.538
4. EMGAAS	0	0	0
5. RESPAM	0	0.086	0.077
6. RESPAS	0.172	0.150	0.161
7. RESPDM	-0.231	-0.293	-0.281
8. RESPDS	0.096	0.131	0.132
9. ECGRAM	0	0	
10. ECGRAS	0.078	0.074	
11. ECGRIM	0	0	
12. ECGRIS	0	0	
13. ECGQRAM	-0.152	-0.138	
14. ECGSRAM	-0.112	-0.141	
15. ECGTRAM	-0.089	0	
16. ECGQDM	0	0	
17. ECGRDM	0	0	
18. ECGSDM	0	0	
19. ECGTDM	0	0	
20. EYESVM	0.115		
21. EYESVS	0		
22. EYEPDM	0.108		
(R^2)	(0.440)	(0.428)	(0.401)

TABLE 8. REGRESSION COEFFICIENTS FOR PREDICTION OF DPSO

Predictor Variables	Variables Included in Regression		
	Physiological and Visual Response Variables	Physiological Variables Only	Selected Physiological Variables
1. EMGHAM	0	0	
2. EMGHAS	0	0	
3. EMGAAM	0.656	0.656	0.631
4. EMGAAS	0	0	0
5. RESPAM	0.136	0.135	0.103
6. RESPAS	0.130	0.146	0.163
7. RESPDM	-0.412	-0.420	-0.386
8. RESPDS	0.163	0.182	0.167
9. ECGRAM	0	0	
10. ECGRAS	0	0	
11. ECGRIM	0	0	
12. ECGRIS	0	0	
13. ECGQRAM	-0.161	-0.135	
14. ECGSRAM	-0.152	-0.162	
15. ECGTRAM	-0.135	0	
16. ECGQDM	0	0	
17. ECGRDM	0	0	
18. ECGSDM	0	0	
19. ECGTDM	0.113	0.088	
20. EYESVM	0.366		
21. EYESVS	-0.286		
22. EYEPDM	0		
(R^2)	(0.417)	(0.399)	(0.375)

SECTION IV

CONCLUSIONS AND RECOMMENDATIONS

RECOMMENDED METRIC

Based on results of this study the following metric for information processing workload (W) is tentatively recommended:

$$W = 0.631\,(EMGAAM) + 0.103\,(RESPAM) + 0.163\,(RESPAS) - 0.386\,(RESPDM) + 0.167\,(RESPDS)$$

Mnemonics are normalized physiological response variables defined as

EMGAAM = Mean arm EMG amplitude

RESPAM = Mean respiration amplitude

RESPAS = Standard deviation of respiration amplitude

RESPDM = Mean respiration duration

RESPDS = Standard deviation of respiration duration

Data requirements to apply the above metric are minimal. Only two recording channels (respiration and arm EMG) are required to provide the necessary raw data. With due consideration to qualifications discussed below, this metric can be applied to estimate relative workload levels in crew-station evaluations of design alternatives. The metric may also have utility for estimating crew/system performance in evaluations where

direct measures of performance are difficult to obtain. Potential for metric application as a performance estimation technique is based on the fact that the above metric and other variations defined in this study were derived as correlates of task performance.

SCALE CHARACTERISTICS

The metric has at least ordinal scale characteristics, and should be interpreted as a relative rather than absolute measure in its present form. In an application of this metric involving comparison of design alternatives A and B, for example, results may indicate that A is preferable to B in terms of expected workload demand on the operator (i.e., the numeric value of W is found to be lower for design A). Quantification of this difference (e.g., "design A produces 20 percent less workload than design B") requires further work on the metric to achieve and demonstrate interval scale characteristics.

Within the study conceptual framework (see Figure 1) workload and performance are assumed to be positively correlated over an intermediate range of task difficulty. This intermediate range is likely to be of interest in evaluating many system configuration trade-offs but the lower range of difficulty would be relevant in other evaluations. The lower "stage I" range in Figure 1 represents the common occurrence of design alternatives that yield no appreciable performance difference but are judged to produce differing workloads. Experimentation to investigate the extremes of task difficulty effects postulated in Figure 1 would aid in refining the above metric into an absolute measure with interval scale characteristics. It

would also demonstrate response of the recommended or other metrics in the stage I range where performance and workload are anticipated to be essentially uncorrelated.

METRIC REFINEMENT AND VALIDATION

Development of an interval scale metric for workload assessment is a desirable ultimate objective. Before that level of development is pursued, the following additional refinement and validation work is needed.

A basic issue is whether the metric should be sensitive to only cognitive load or to information processing workload in a more generic sense. Piloting and other flight-crew tasks in an operational environment impose composites of stress and perceptual, cognitive, and motor response demands. The current metric, for example, includes one predictor variable (arm EMG) sensitive to arm motions required for aircraft control or switch activation. These tasks could be viewed as an integral part of the crew member's information processing activity. Other physiological response variables not presently included in the metric (e.g., galvanic skin response and skin impedance) are sensitive to stress.

In the present study, the secondary task was applied as an independent variable to increase overall task loading. A viable approach to supplement physiological measures is to apply a similar task as an indicator of reserve capacity (i.e., as a dependent variable). The most effective way to combine various types of dependent measures such as physiological, opinion, and secondary task responses remains to be determined. An alternative

approach to that applied in the present study is to develop a test battery of workload-related measures with elements that can be applied and evaluated individually or collectively.

Criterion variance accounted for by predictor variables analyzed in this study was nominally $R^2 = 0.4$. One approach to obtaining larger values of R^2 may be to include additional physiological and secondary task variables as predictors. A second approach would be application of various data transformations to predictor variables in an attempt to achieve higher correlations. Evaluation of the possible benefits of applying transformations prior to regression analysis was not within the current study scope.

Refinement techniques noted above have potential for improving the workload metric recommended. However, to establish the metric's general utility for crew workload assessment, its validity must be demonstrated in a variety of real-world situations. Validation efforts on more complete mission-task simulators are needed, followed by in-flight verification.

RECOMMENDATIONS FOR FURTHER STUDY

The following steps are recommended for metric refinement and validation:

1. Additional analysis of various combinations of physiological variables, including evaluation of alternative raw data transformations.

2. Additional analysis of other physiological response variables (e.g., skin impedance) and other features of currently measured responses.

3. Evaluation of secondary tasks (e. g., time estimation) as measures of reserve capacity, and approaches to combining secondary task and physiological measures (e. g., regression context versus test battery).

4. Validation of recommended or refined workload metrics in real-world task environments.

APPENDIX A

INSTRUCTIONS TO PILOTS

APPENDIX A

INSTRUCTIONS TO PILOTS

GENERAL INFORMATION

You will be asked to perform two simultaneous tasks. One is a flight control task simulating a landing approach on instruments. The other is a secondary task which places additional demands on your attention. This secondary task may be considered as an approximation of the demands on your attention caused by various tasks other than flight control under certain operational conditions. We will be recording data indicating how well you are able to perform each of these tasks simultaneously. It is important that you consistently perform both tasks as well as possible.

FLIGHT CONTROL TASK

The simulated aircraft you will be flying approximates the flight characteristics of an F-4. The aircraft is initially trimmed for level flight at 165 knots, on the desired flight path, and 6.5 miles from the runway. During the simulated approach, maintain the desired flight path as closely as possible by maneuvering the aircraft to keep the path error symbol centered on the aircraft symbol. Also, attempt to maintain approach speed at 165 knots. We will be recording deviations from desired flight path and approach speed. The simulated approach will automatically be terminated at an altitude of 200 feet. Approximate flight time to this point is two minutes.

The above instructions apply to all flights. Difficulty of the flight control task will be changed from one flight to another by varying turbulence conditions and flight control system characteristics.

SECONDARY TASK

During a flight, you will hear letters from the alphabet spoken through headsets. Your task will be to push the pitch trim switch in the forward (pitch down) direction after hearing certain letters, and in the aft (pitch up) direction after hearing all other letters. The trim switch is not connected to the aircraft, and has no affect on trim control.

On some flights, you will be instructed to push the switch forward only when you hear one of the two letters, A or H, and aft when you hear any other letter.

On other flights, you will be instructed to push the switch forward only when you hear one of the four letters A, H, J, or Q, and aft when you hear any other letter.

Your response time to these letters will be recorded. The computer will also record whether you pushed the switch in the correct direction. Attempt to respond as quickly as possible each time a letter is presented, and to keep response errors to a minimum.

A reminder of which secondary task response is required will be displayed above the simulated flight instruments prior to and during each flight. Display of the number "2" indicates that you should push the switch forward

for the two letters A and H, and aft for other letters. Display of the number "4" indicates that you should push the switch forward for the four letters A, H, J, and Q, and aft for other letters.

APPENDIX B

SELECTED PLOTS OF NORMALIZED RESPONSE DATA

APPENDIX B
LIST OF ILLUSTRATIONS

APPENDIX B
LIST OF ILLUSTRATIONS (concluded)

APPENDIX B

SELECTED PLOTS OF NORMALIZED RESPONSE DATA

The following illustrations are those identified in Table 2, Section III, as containing statistically significant ($p < 0.05$) main effects or interactions. Test condition numbers are defined in Figure 2, Section II. Plots indicate means of normalized (Z-score) response data, and one standard-deviation ranges around the means.

Scores in the +Z direction correspond to larger values of the raw-score measure (e.g., longer durations and response times, larger errors, and increased task-difficulty ratings).

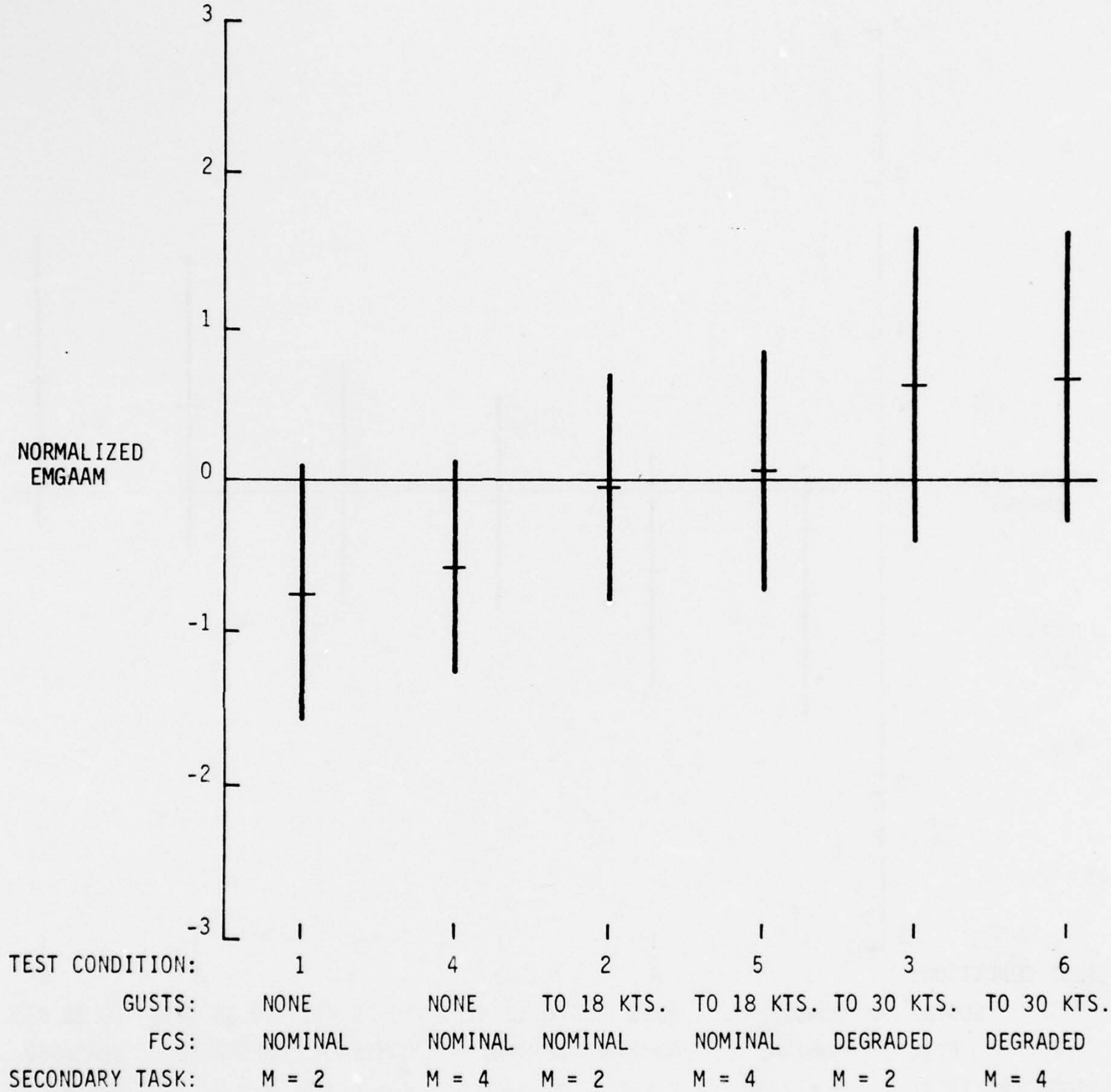


Figure B1. Normalized Mean Arm EMG Amplitude (EMGAAM) versus Test Condition

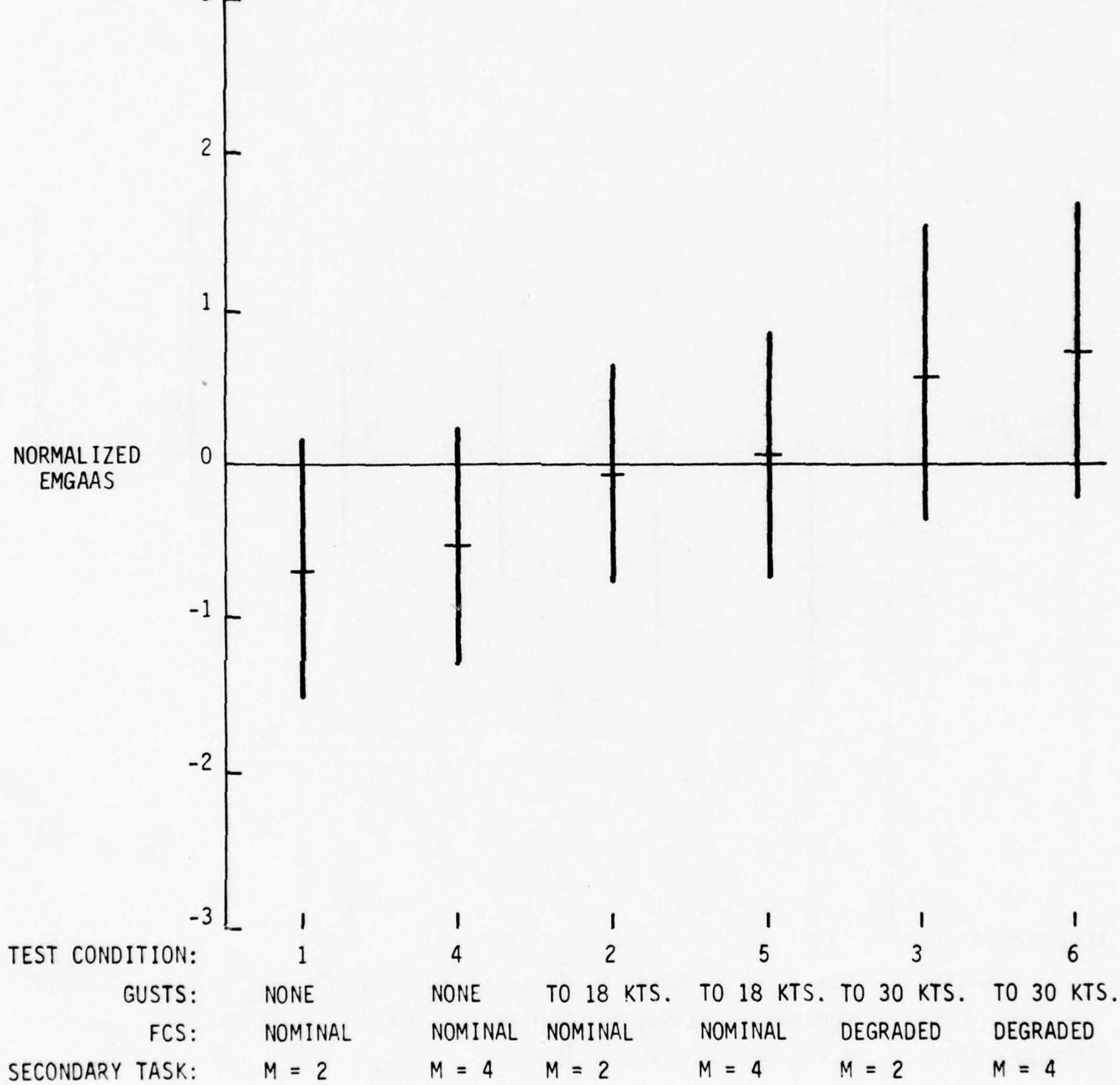


Figure B2. Normalized Standard Deviation Arm EMG Amplitude (EMGAAS) versus Test Condition

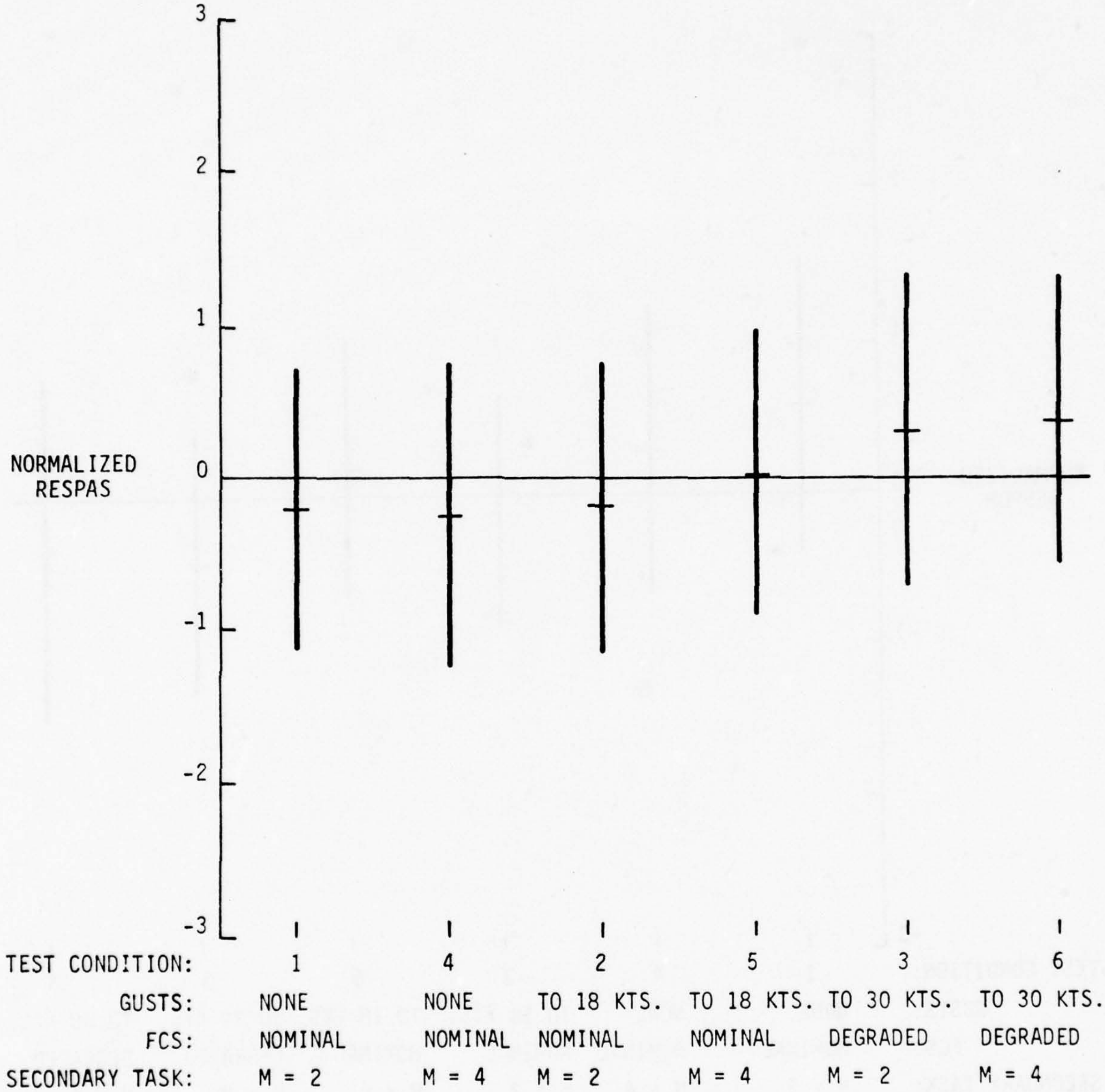


Figure B3. Normalized Standard Deviation Respiration Amplitude (RESPAS) versus Test Condition

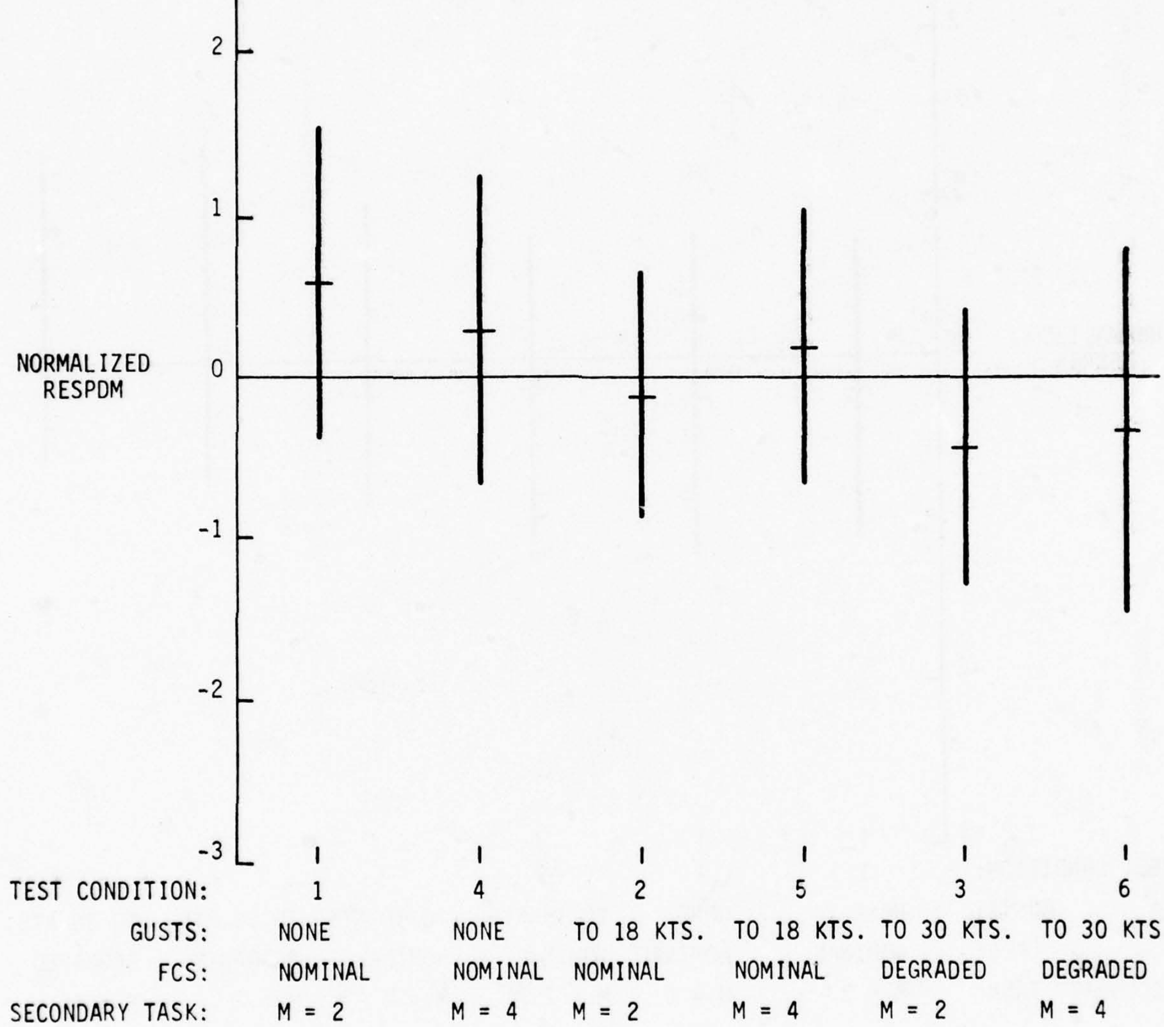


Figure B4. Normalized Mean Respiration Duration (RESPDM) versus Test Condition

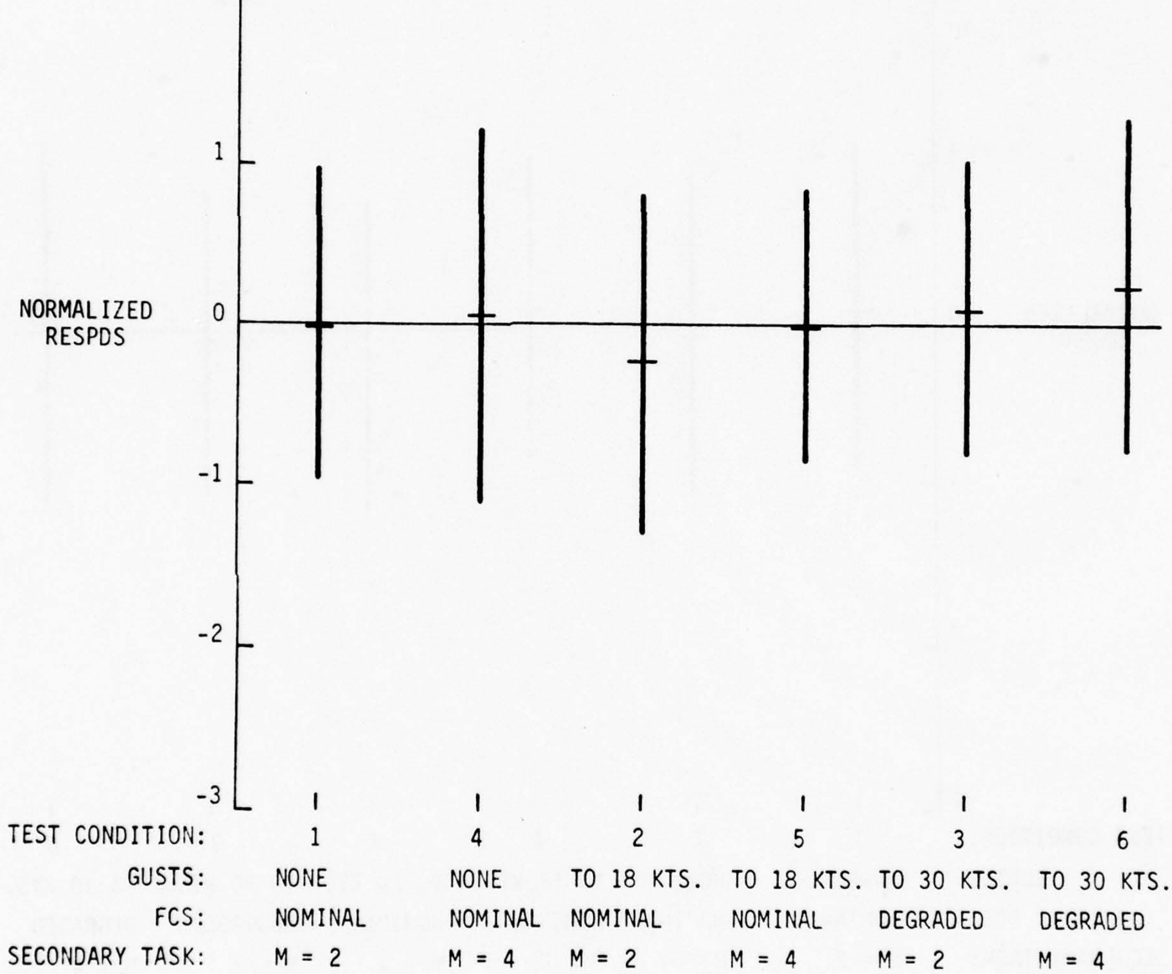


Figure B5. Normalized Standard Deviation Respiration Duration (RESPDS) versus Test Condition

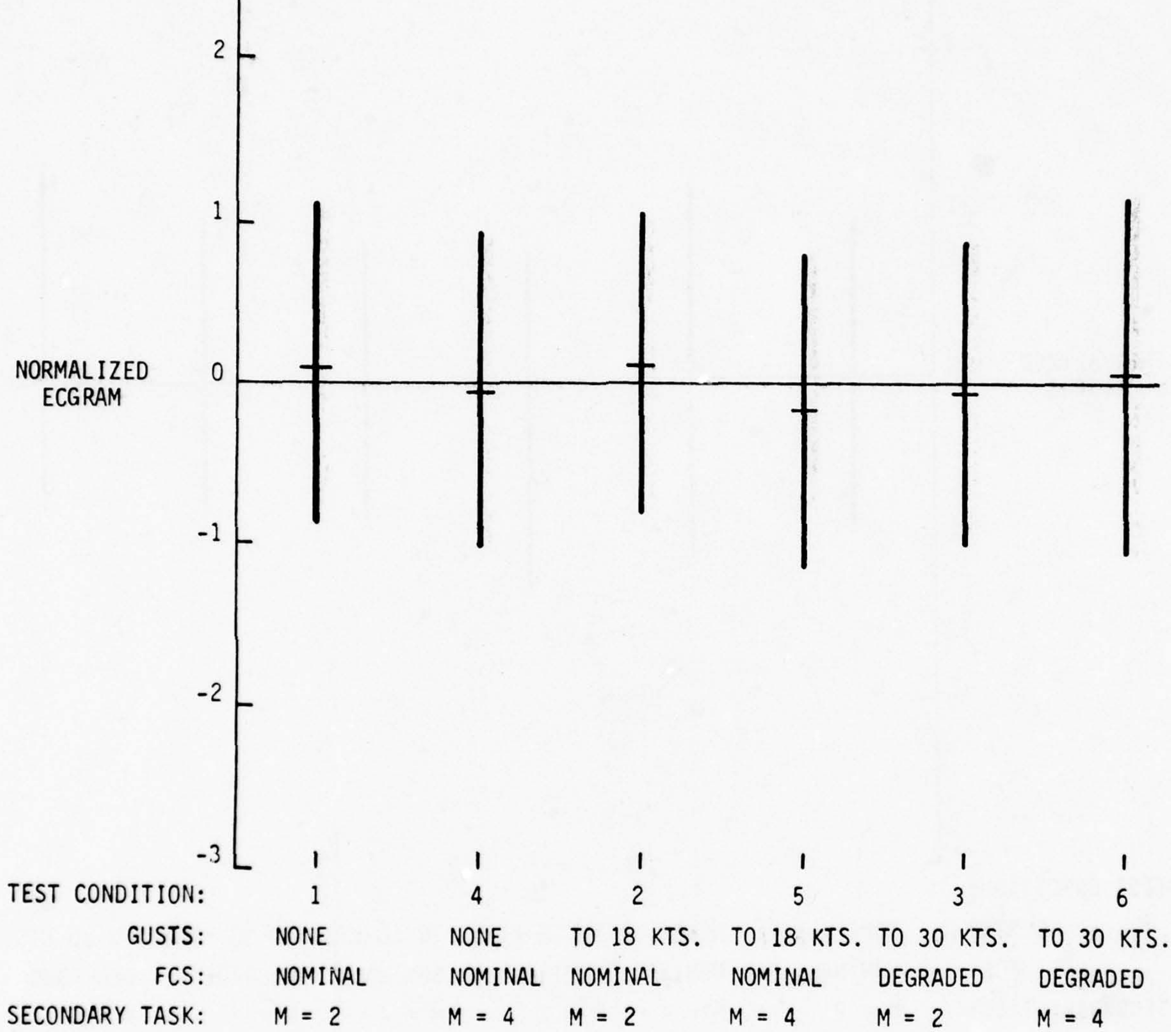


Figure B6. Normalized Mean ECG R-wave Amplitude (ECGRAM) versus Test Condition

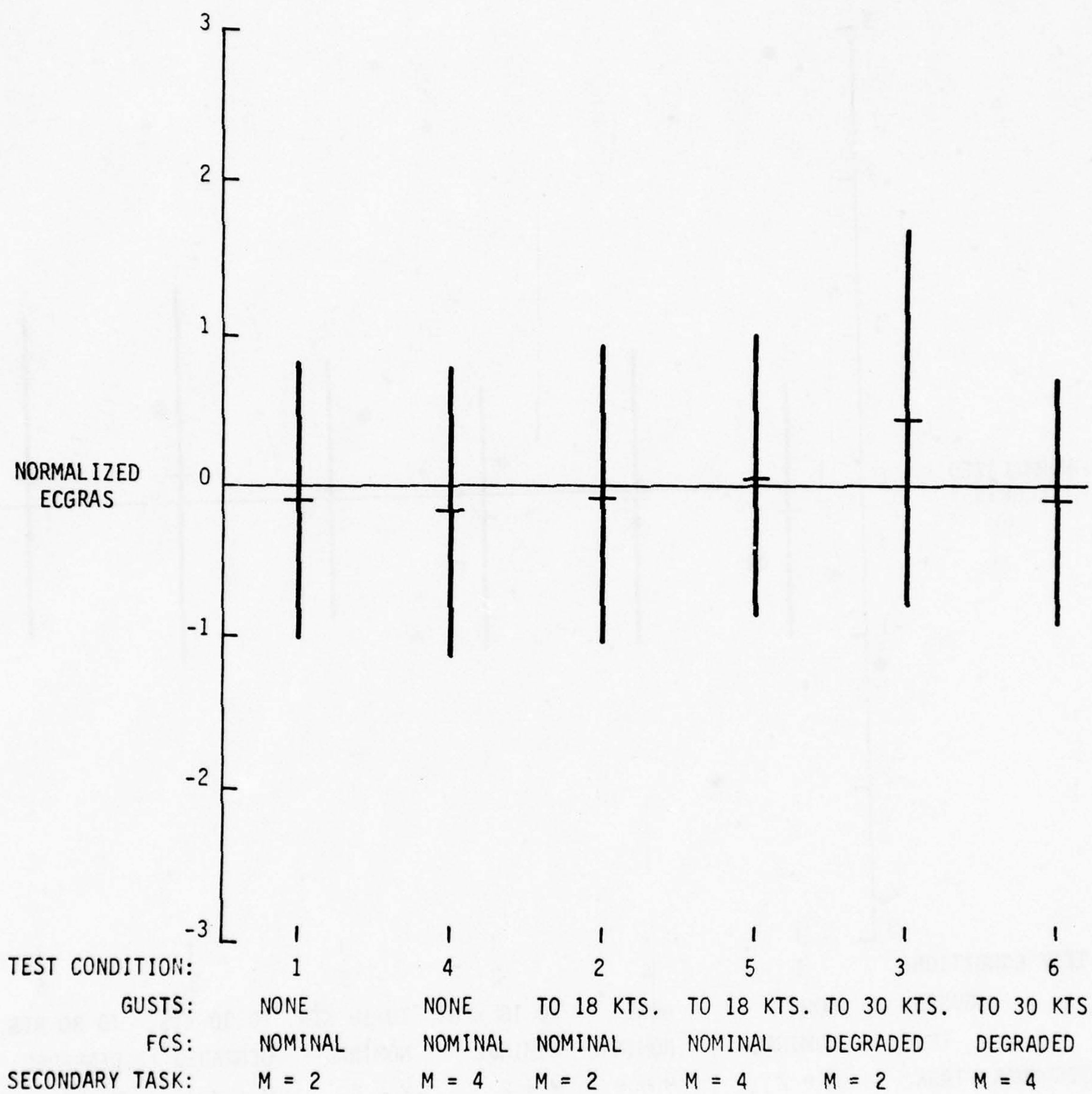


Figure B7. Normalized Standard Deviation ECG R-wave Amplitude (ECGRAS) versus Test Condition

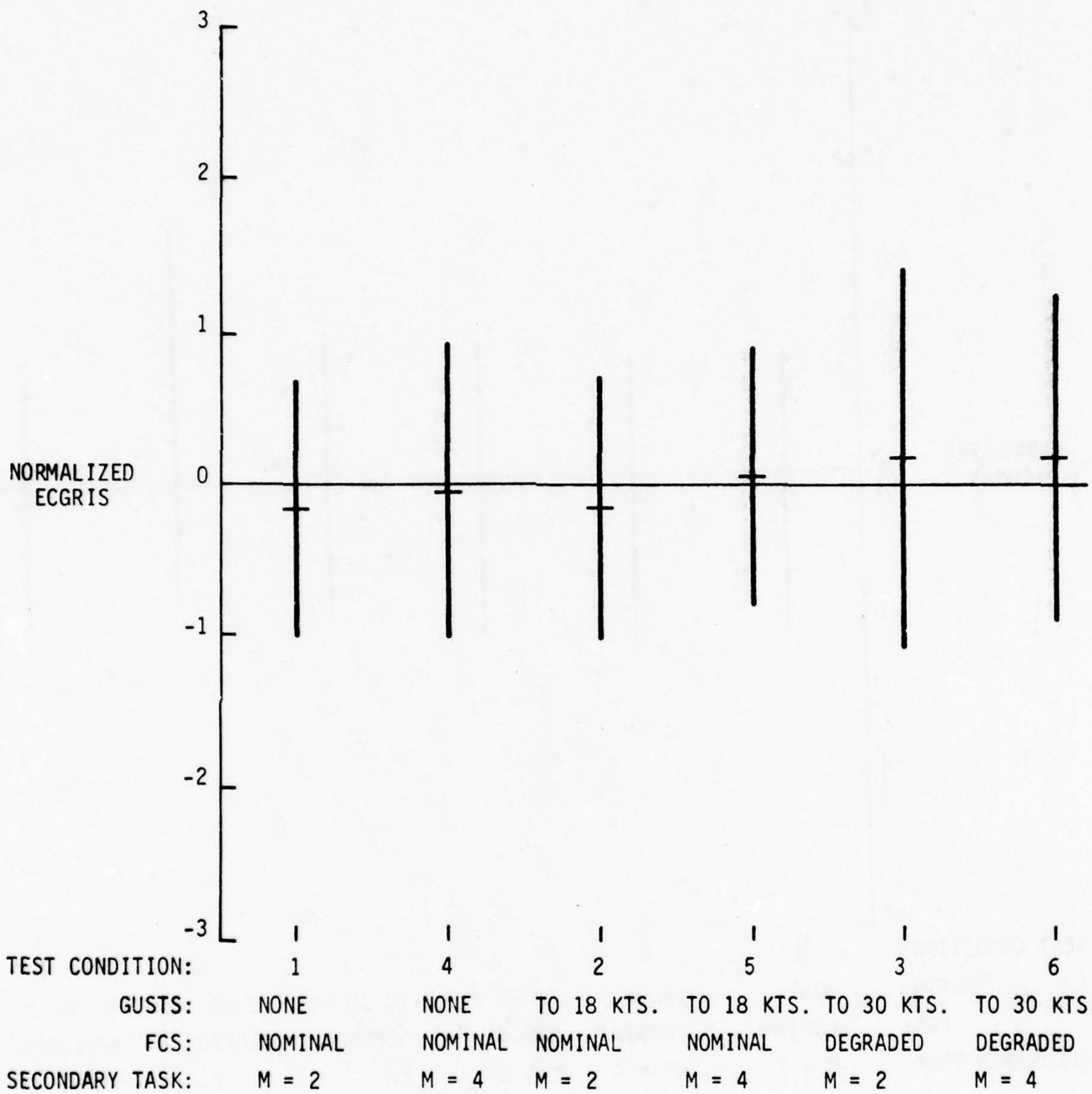


Figure B8. Normalized Standard Deviation ECG R-wave Interval (ECGRIS) versus Test Condition

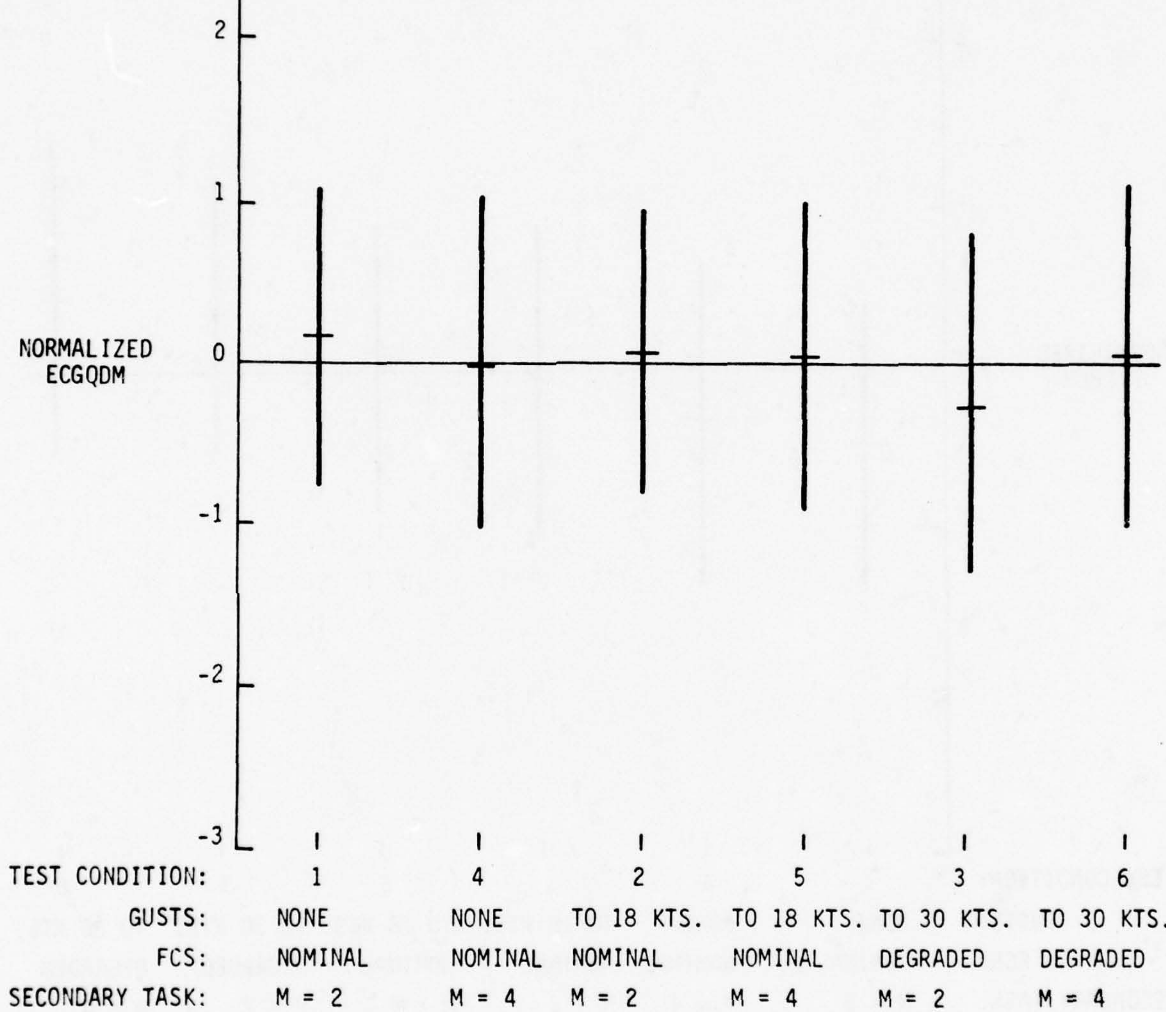


Figure B9. Normalized Mean ECG Q-wave Duration (ECGQDM) versus Test Condition

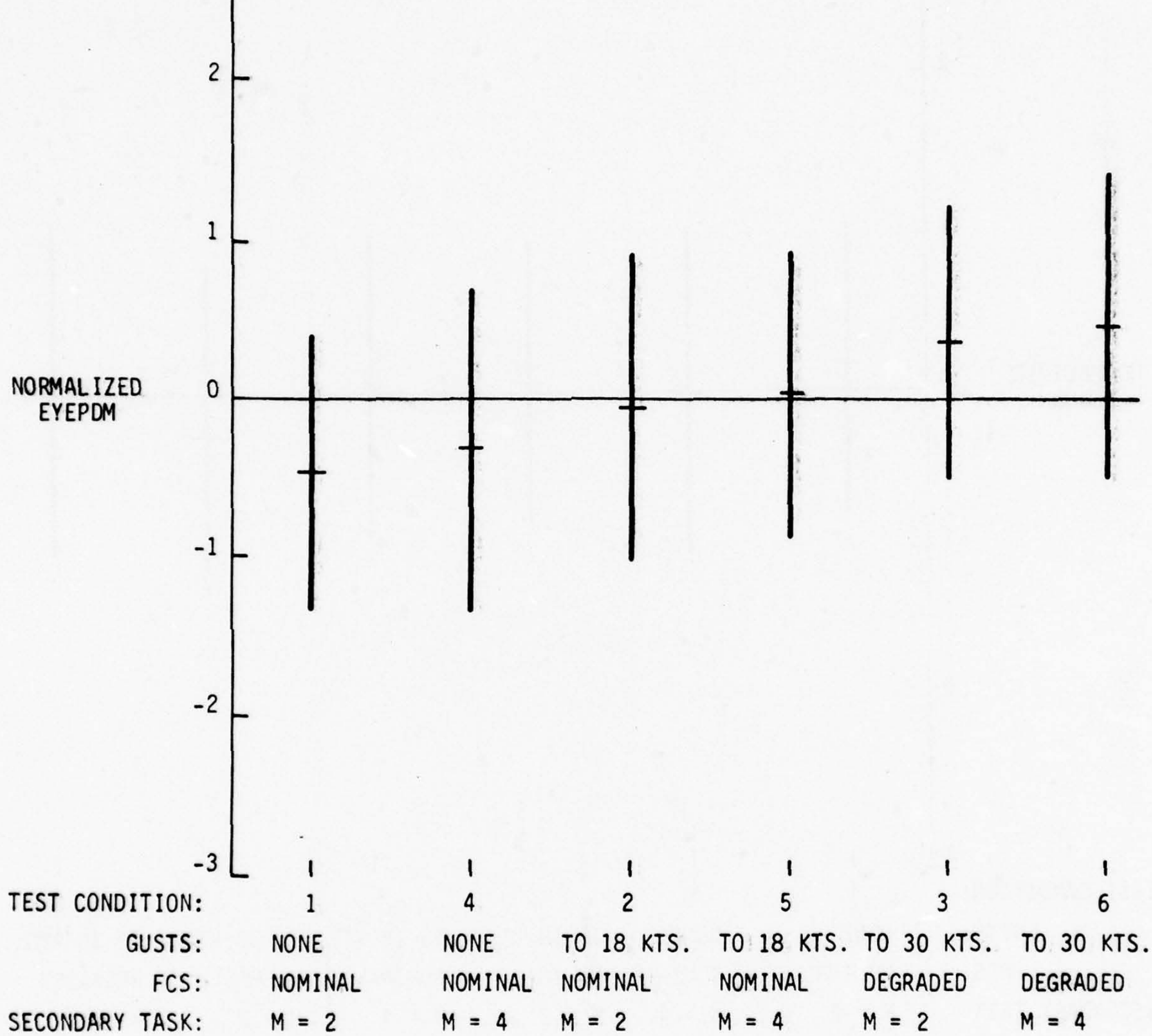


Figure B10. Normalized Mean Eye Pupil Diameter (EYEPDM) versus Test Condition

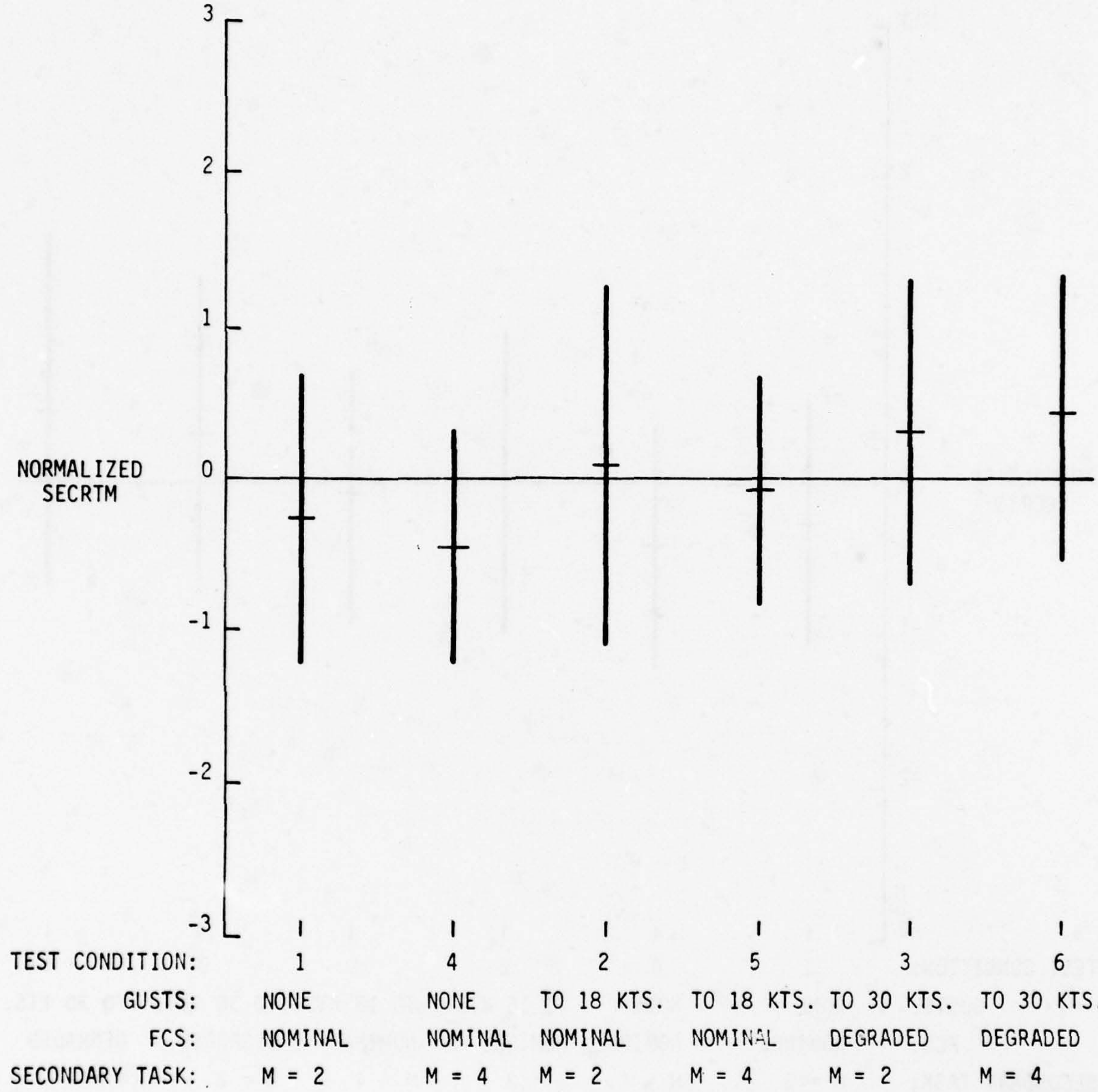


Figure B11. Normalized Mean Secondary Task Response Time (SECRTM) versus Test Condition

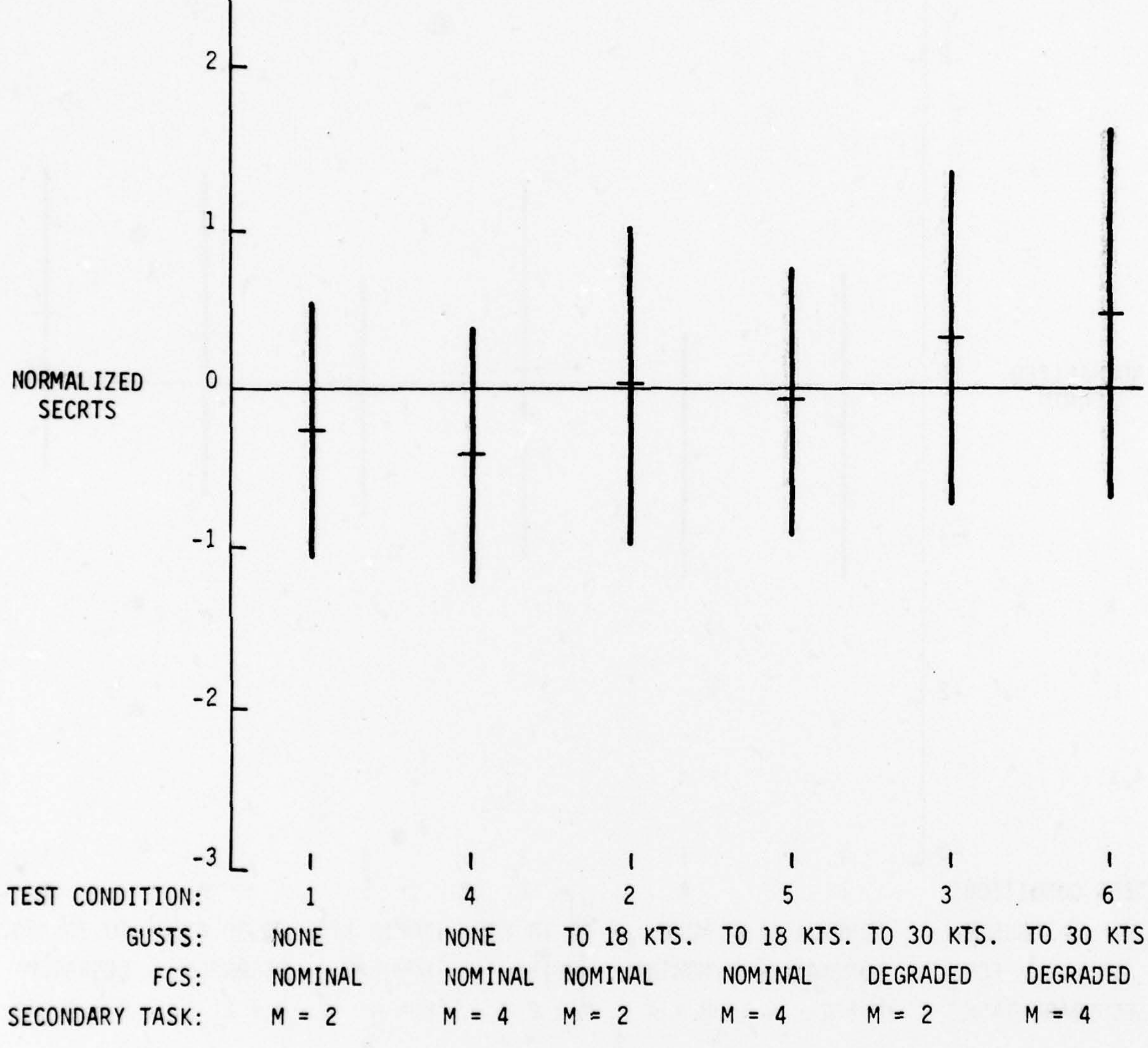


Figure B12. Normalized Standard Deviation Secondary Task Response Time (SECRTS) versus Test Condition

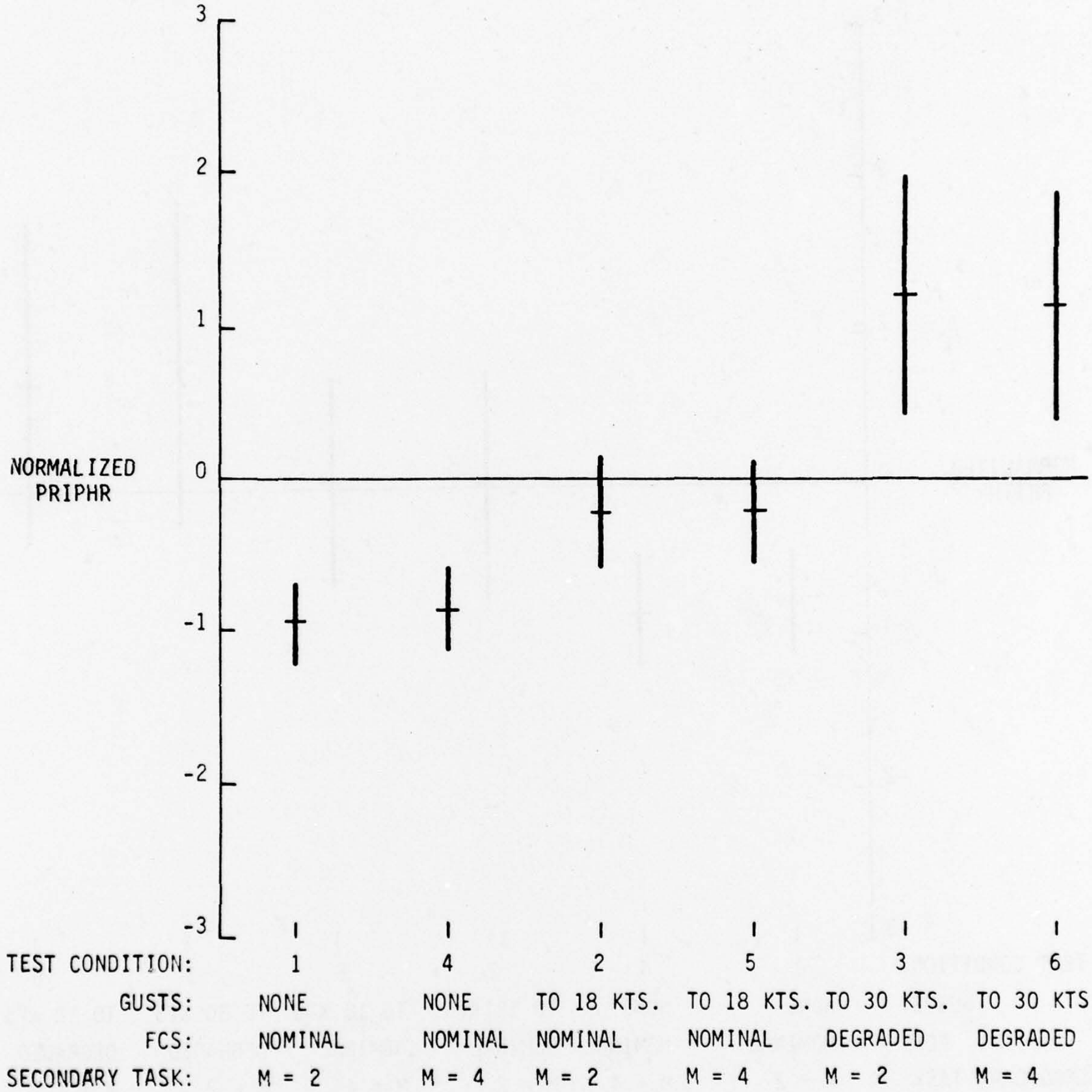


Figure B13. Normalized RMS Primary Task Roll Attitude (PRIPHR) versus Test Condition

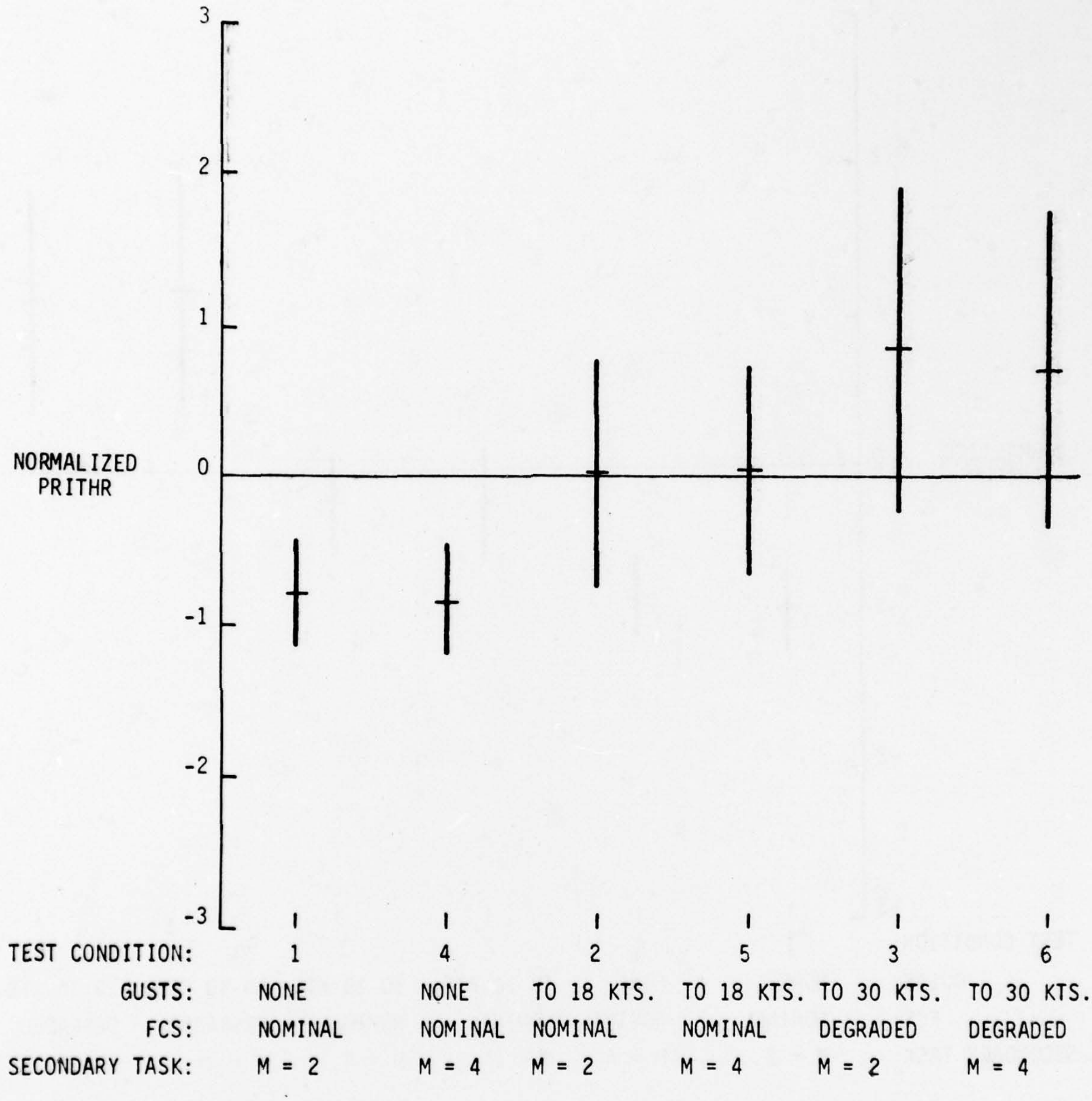


Figure B14. Normalized RMS Primary Task Pitch Attitude (PRITHR) versus Test Condition

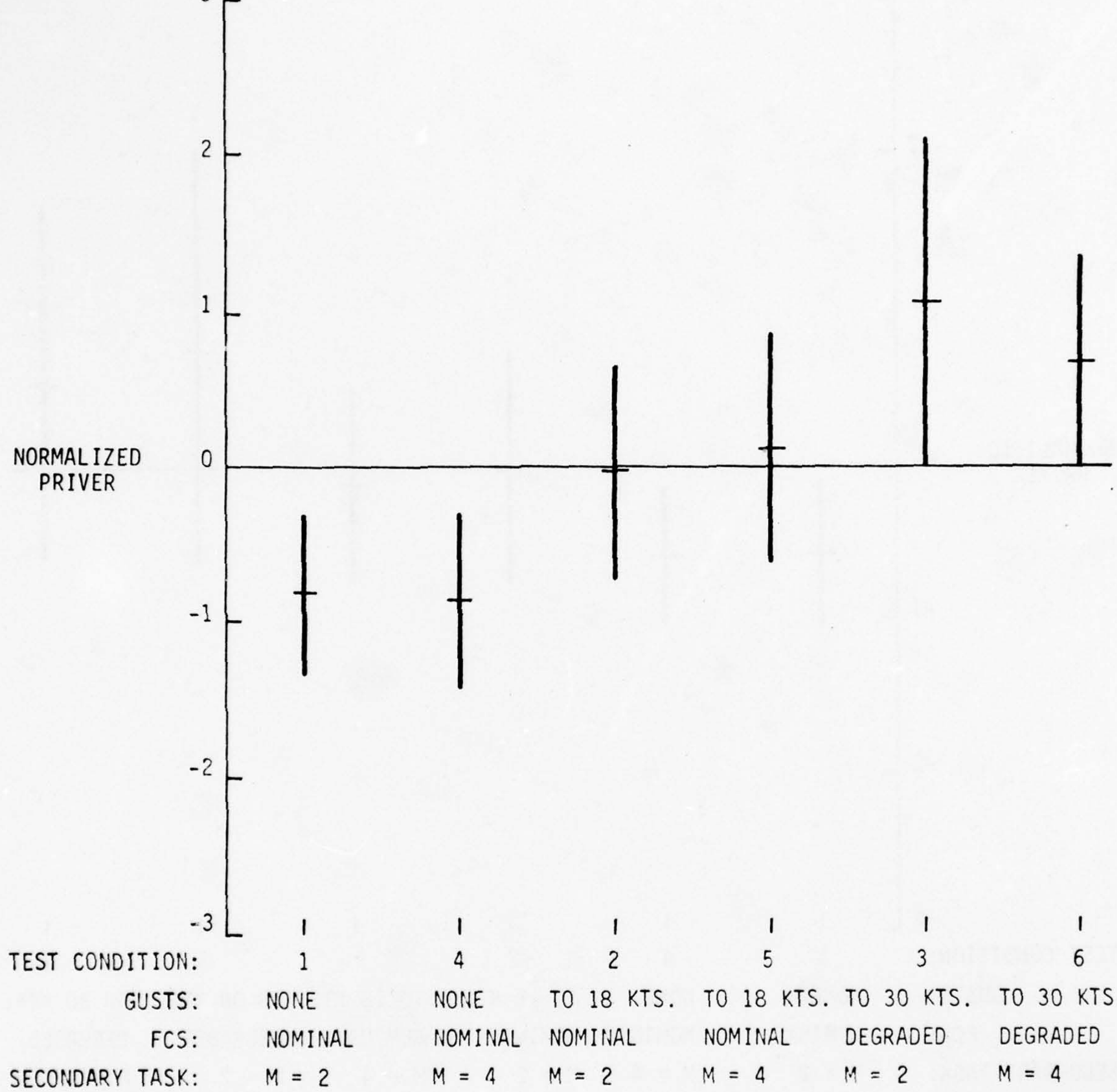


Figure B15. Normalized RMS Primary Task Speed Error (PRIVER) versus Test Condition

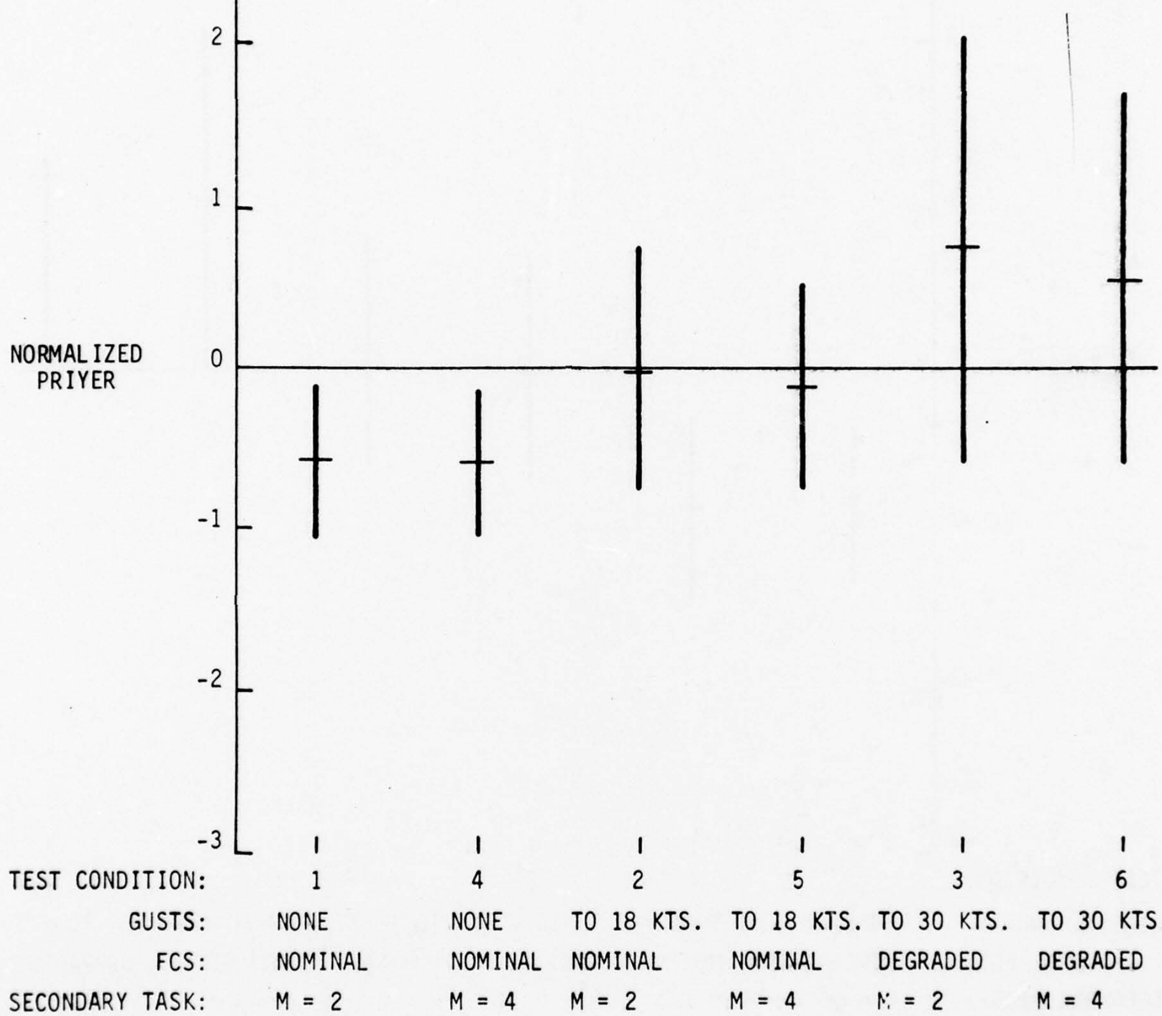


Figure B16. Normalized RMS Primary Task Lateral Path Error (PRIYER) versus Test Condition

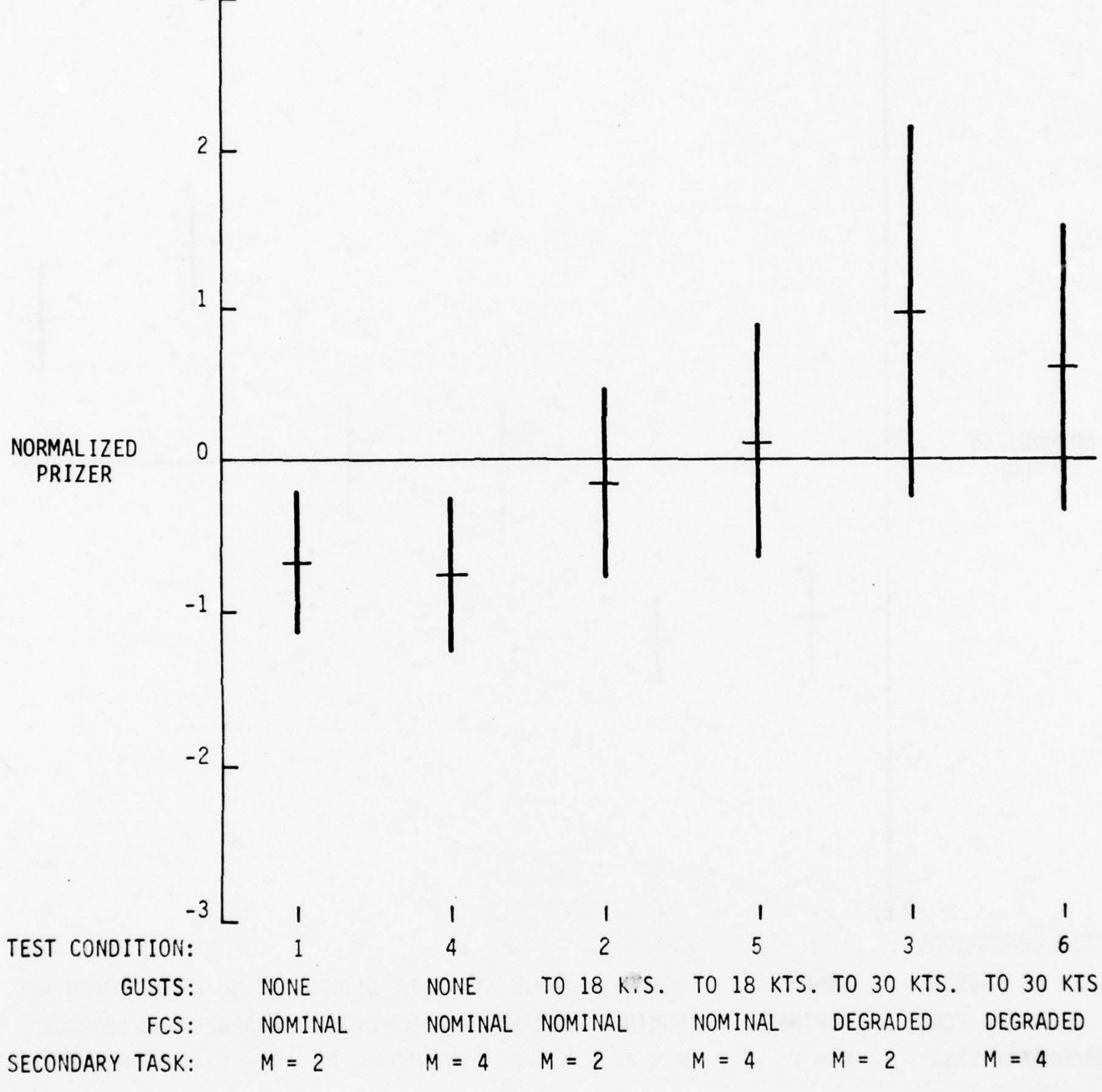


Figure B17. Normalized RMS Primary Task Vertical Path Error (PRIZER) versus Test Condition

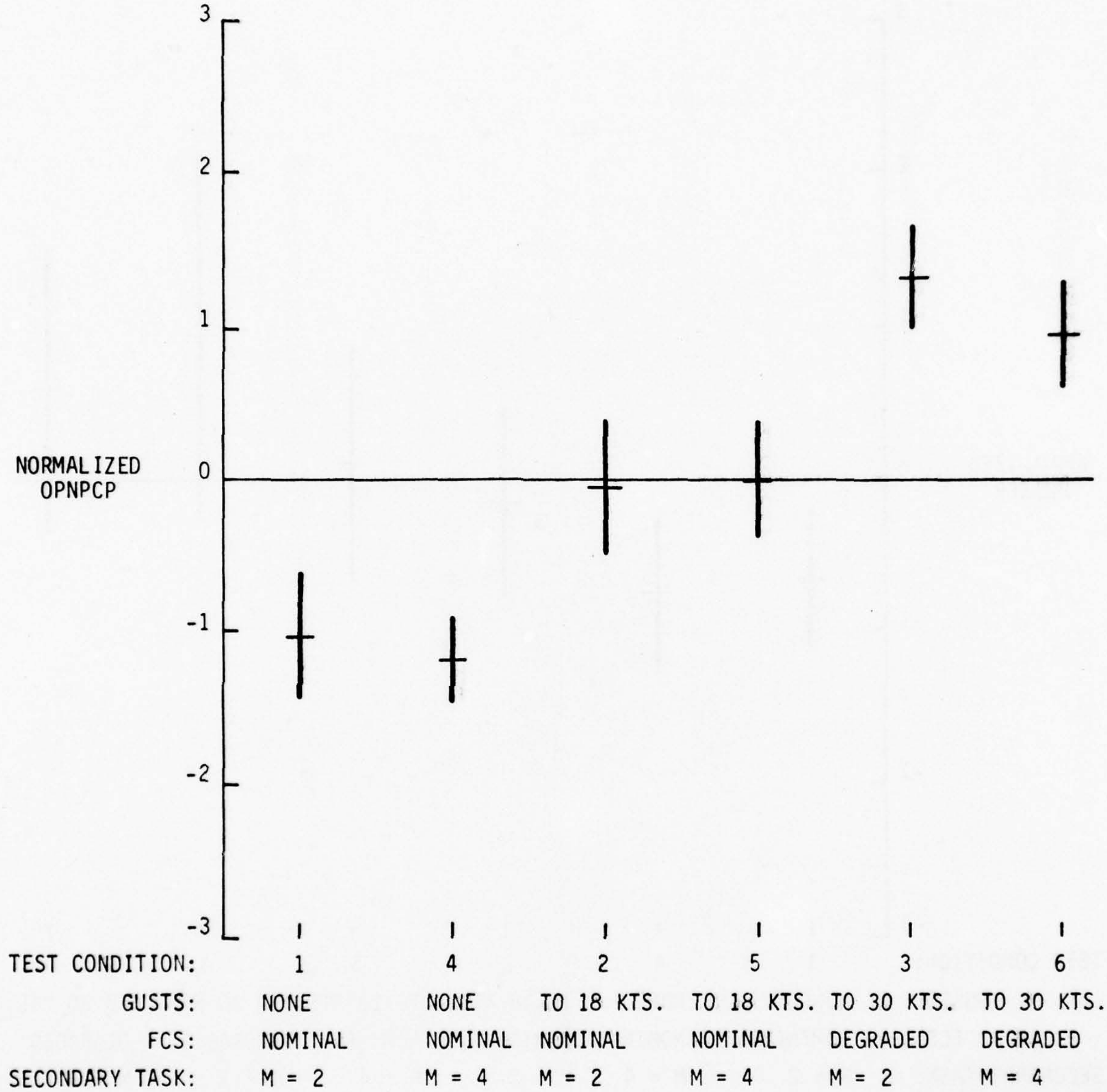


Figure B18. Normalized Proportion More-difficult Judgments, Paired Comparison Opinion (OPNPCP) versus Test Condition

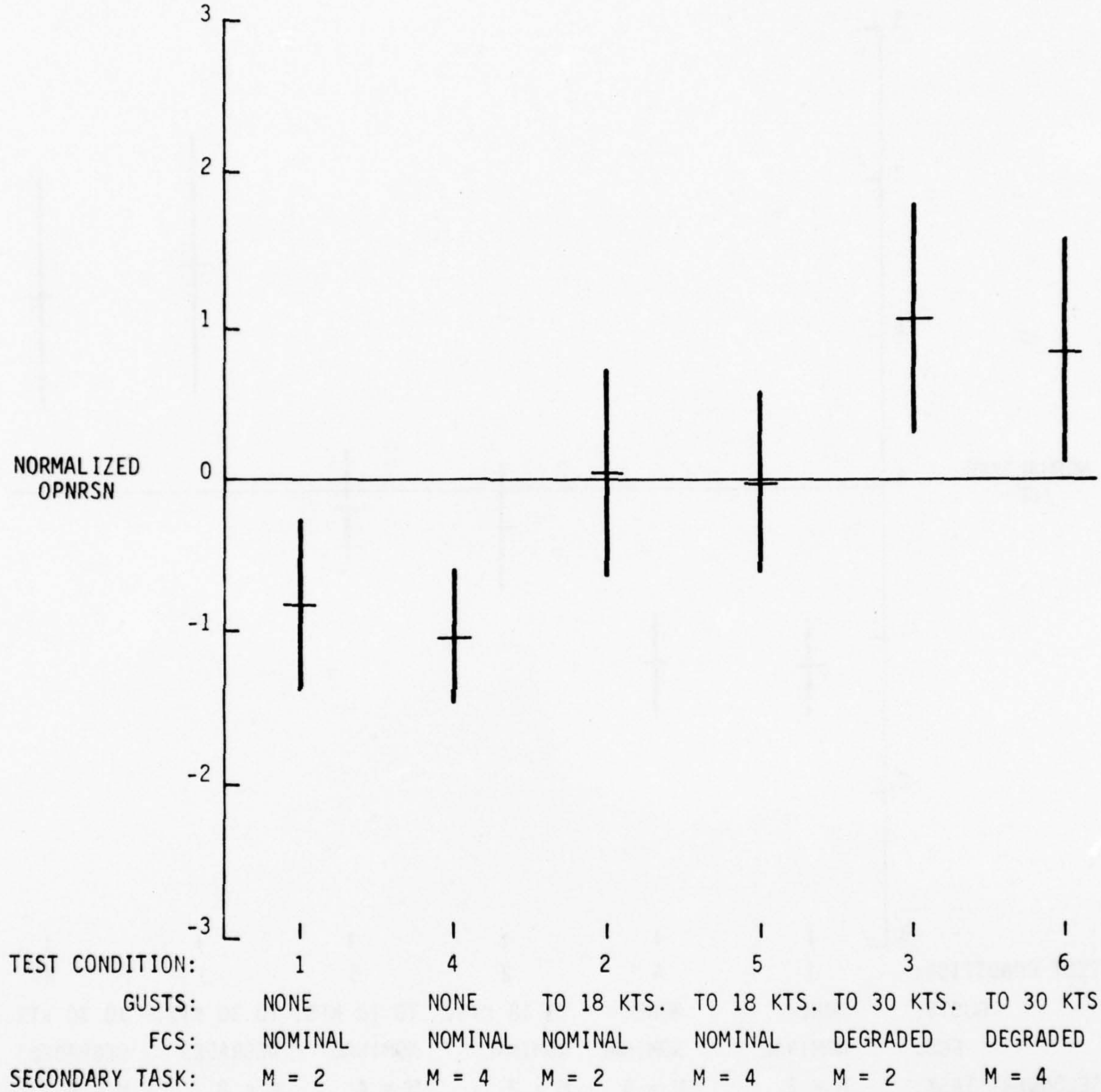


Figure B19. Normalized Numeric Rating, Rating-Scale Opinion (OPNRSN) versus Test Condition

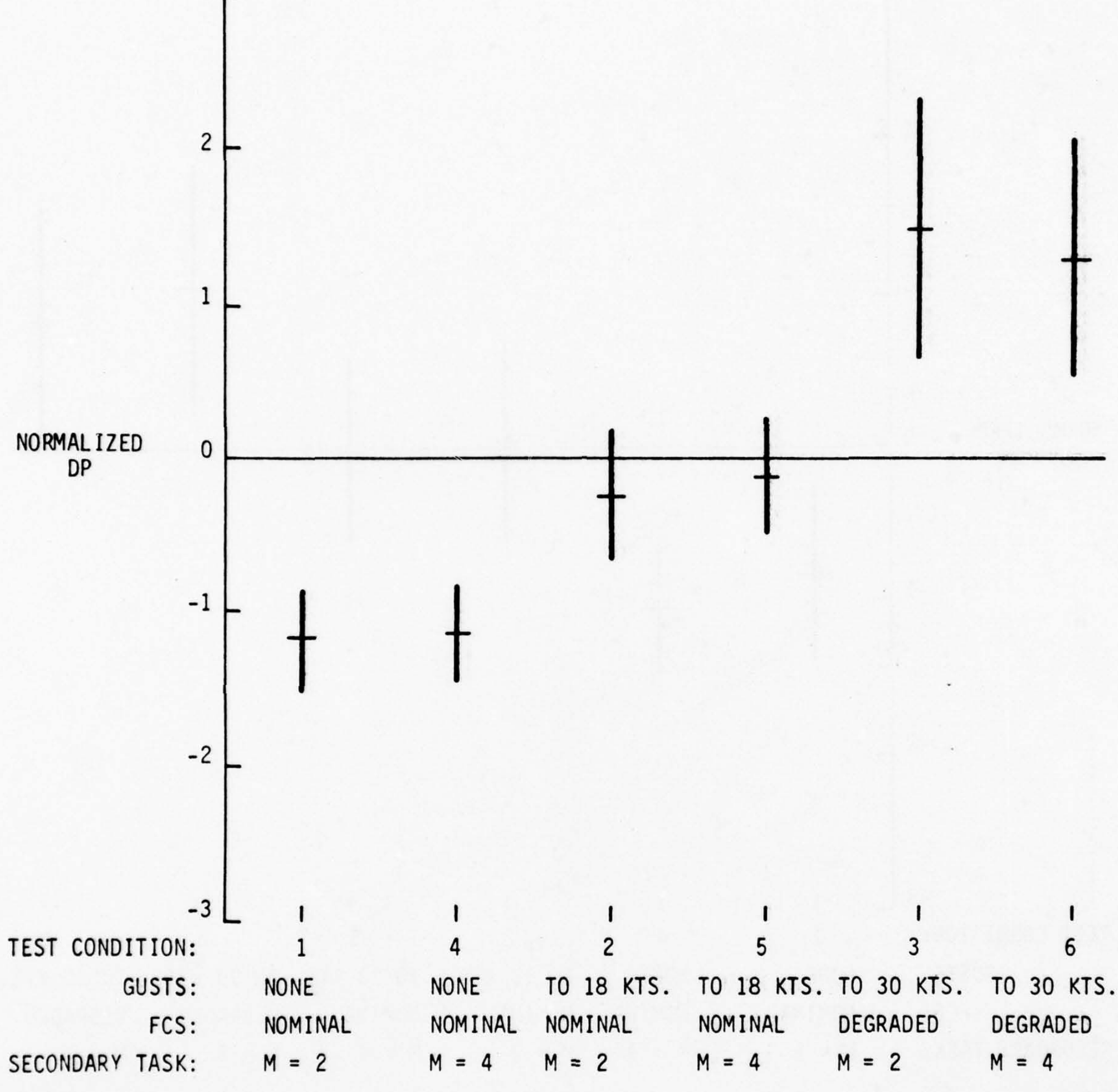


Figure B20. Normalized Discriminant Scale Response, Primary Task Measures Only (DP) versus Test Condition

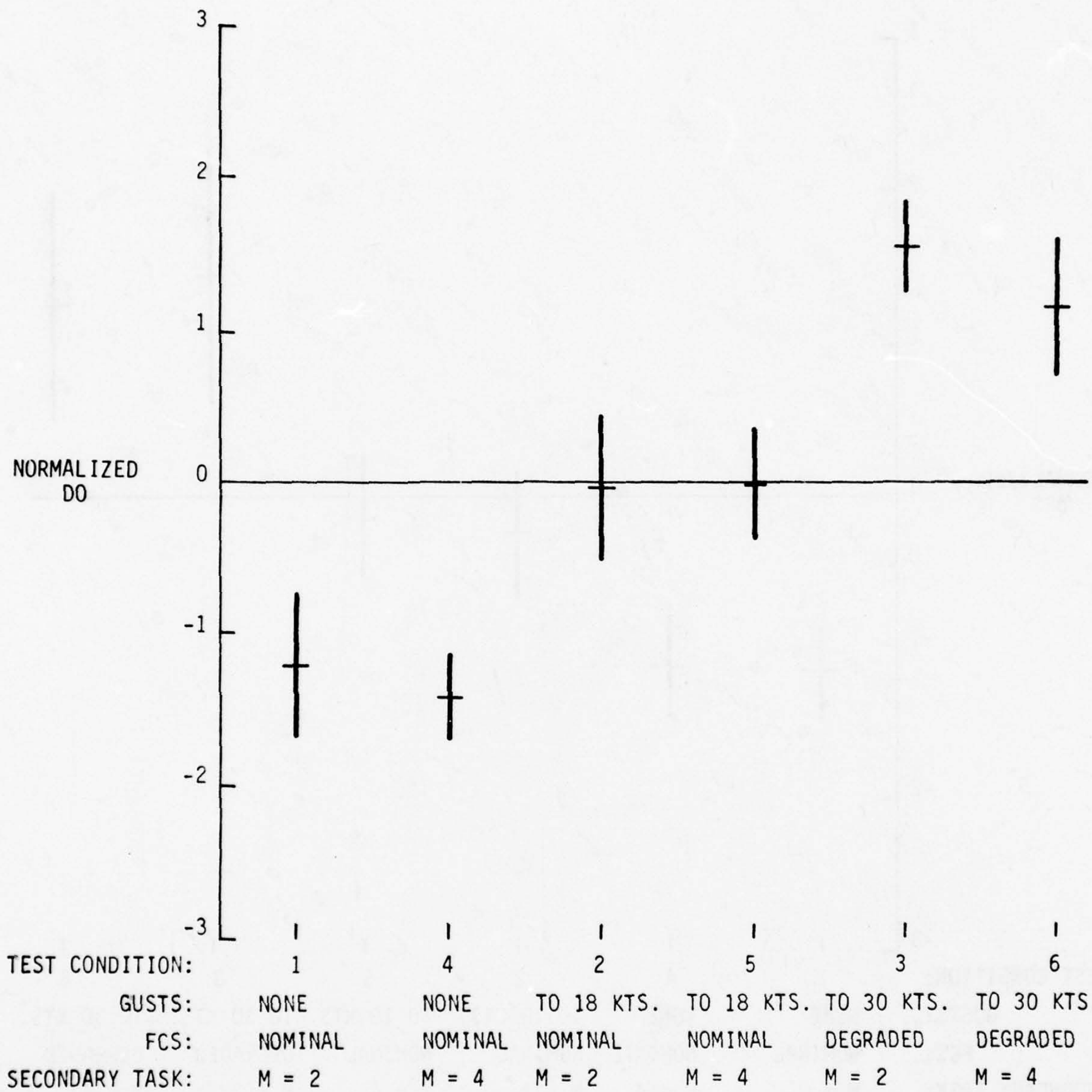


Figure B21. Normalized Discriminant Scale Response, Opinion Measures Only (DO) versus Test Condition

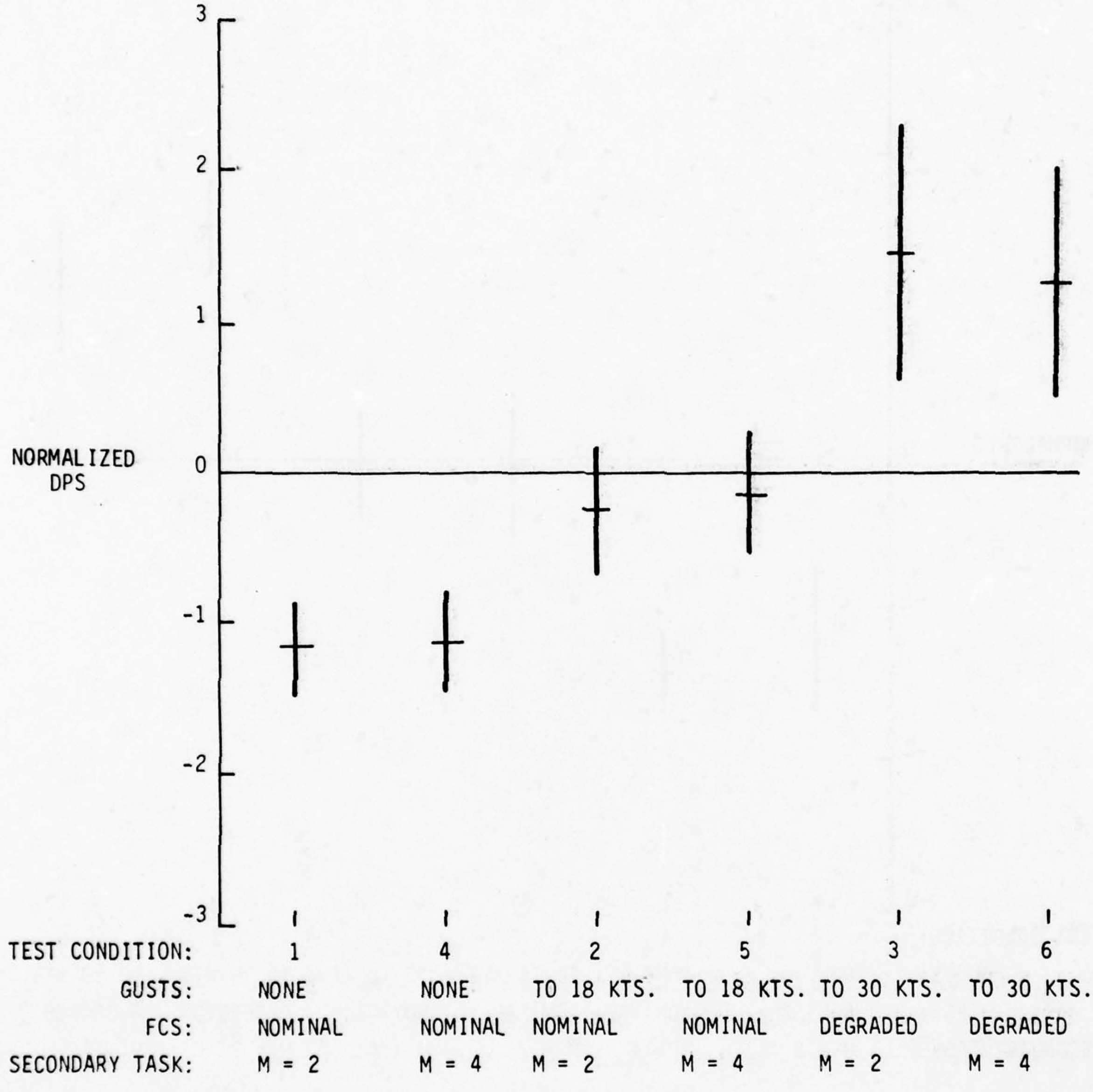


Figure B22. Normalized Discriminant Scale Response, Combined Primary and Secondary Task Measures (DPS) versus Test Condition

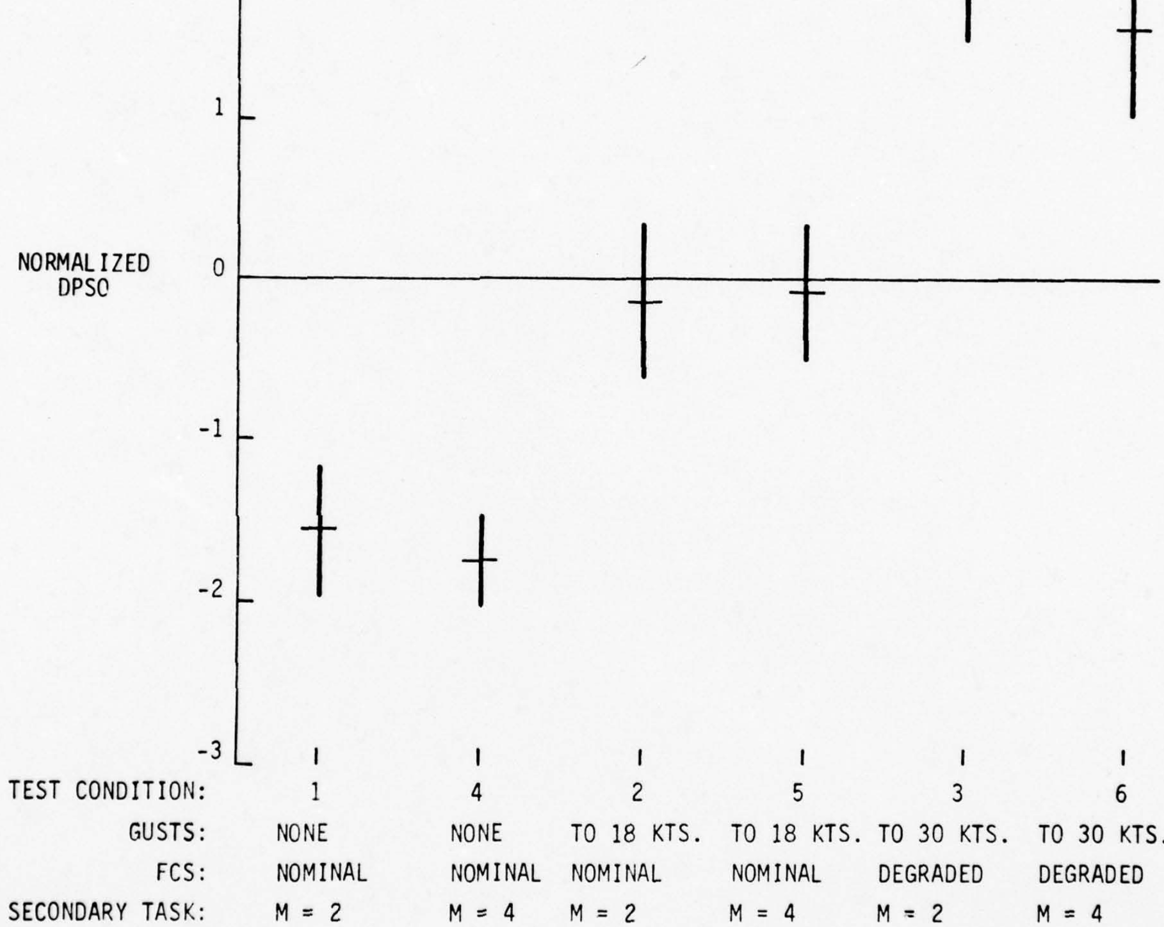


Figure B23. Normalized Discriminant Scale Response, Combined Primary Task, Secondary Task, and Opinion Measures (DPSO) versus Test Condition

REFERENCES

1. D.A. Spyker, S.P. Stackhouse, A.S. Khallafalla, and R.C. McLane, Development of Techniques for Measuring Pilot Workload, National Aeronautics and Space Administration, Washington, D.C. 20546, NASA CR-1888, November 1971.

2. S.P. Stackhouse, Workload Evaluation of LLNO Display, 7201-3408, Honeywell Inc., Minneapolis, Minnesota, October, 1973.

3. S.P. Stackhouse, Pilot Workload Assessment, F0289-FR1, Honeywell Inc., Minneapolis, Minnesota, September, 1975.

4. S.P. Stackhouse, Measurement of Pilot Workload in Manual Control Systems, F0398-FR1, Honeywell Inc., Minneapolis, Minnesota, January 1976.

5. S. Sternberg, "Memory-Scanning: Mental Process Revealed by Reaction-time Experiments," American Scientist, Vol. 57, 1969, pp. 421-457.

6. J.D. Wolf, MLS Information Display, Project F4031, Honeywell Inc., Minneapolis, Minnesota, December, 1974.

7. M.J. Krebs, J.W. Wingert, and T. Cunningham, Exploration of an Oculometer-Based Model of Pilot Workload, Report No. 76SRC39, Honeywell Inc., Minneapolis, Minnesota, March, 1977.

8. H.M. Simpson and S.M. Hale, "Pupillary Changes During a Decision Making Task," Perceptional and Motor Skills, Vol. 29, pp. 495-498.

9. J.L. Bradshaw, "Load and Pupillary Changes in Continuous Processing Tasks," British Journal of Psychology, Vol. 59, 1968, pp. 265-271.

10. F-4 Aerodynamic Derivatives, 9842, McDonnell Aircraft Company, St. Louis, Missouri.

11. R.K. Heffley and W.F. Jewell, Aircraft Handling Qualities Data, 1004-1, Systems Technology Inc., Hawthorne, California, May, 1972.

12. G.E. Cooper and R.P. Harper, The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities, National Aeronautics and Space Administration, Washington, D.C. 20546, NASA TN D-5153, April, 1969.

13. W.W. Cooley and P.H. Lohnes, Multivariate Procedures for the Behavioral Sciences, New York: Wiley and Sons, 1962.

14. N.R. Draper and H. Smith, Applied Regression Analysis, New York: Wiley and Sons, 1966.

15. M.M. Tatsuoka, Multivariate Analysis: Techniques for Educational and Psychological Research, New York: Wiley and Sons, 1971.